ROYAL SOCIETY
OF **CHEMISTRY**

PAPER
María Victoria Gil, Berend Smit *et al.*
Biomass to energy: a machine learning model
for optimum gasification pathways

Biomass to energy: a machine learning model for optimum gasification pathways†

María Victoria Gil, [ID] *[a] Kevin Maik Jablonka, [ID] [b] Susana Garcia,[c] Covadonga Pevida[a] and Berend Smit [ID] *[b]

Biomass is a highly versatile renewable resource for decarbonizing energy systems. Gasification is a promising conversion technology that can transform biomass into multiple energy carriers to produce heat, electricity, biofuels, or chemicals. At present, identifying the best gasification route for a given biomass relies on trial and error, which involves time-consuming experimentation that, given the wide range of biomass feedstocks available, slows down the deployment of the technology. Here, we use a supervised non-parametric machine-learning method, Gaussian process regression (GPR), that provides robust predictions when working with small datasets, to develop a model to find the optimal application of a particular biomass in gasification processes. Leave-one-out cross-validation (LOOCV) is used to validate the model's predictive performance. Our model can select the suitable gasification pathway from the characteristics of the biomass, and also identify the optimal operating conditions for a selected application of the produced gas. In addition, with this model, we can obtain insights into the relationships between biomass properties and gasification results, leading to a better understanding of the process. A relevant aspect of this work is that these results rely on a relatively small dataset, representative of those typically collected by research groups using different types of gasifiers worldwide. This study opens the path for future integration of such data, which would allow addressing the complexity of biomass and conversion process simultaneously. With this work, we aim to increase the flexibility of biomass gasification processes and promote the development of bioenergy technologies, considered crucial in the energy transition context.

Introduction

Bioenergy is one of the key pillars to decarbonize our global energy systems, and is estimated to contribute 17–20% to the total energy supply by 2050 in proposed net-zero scenarios.[1,2] The importance of bioenergy for the energy transition has been increasingly underlined due to its versatility in substituting fossil fuels and the possibility of generating negative emissions.[3] Additionally, biomass is very appealing because of its diversity, local availability, immense commercialization potential, carbon-neutral nature, and renewable characteristics.[4–9] This is highlighted by the European Commission in the European Green Deal, where the biomass' potential to provide

a solution that delivers renewable energy and negative emissions, together with sustainably managed forests and feedstock sources, is recognized.[10] Within this context, biomass utilization needs to focus on sustainable biomass feedstocks with minimal impact on food security and biodiversity. Examples include biomass-based wastes (e.g., industrial or agroforestry wastes), which will be key for bioenergy and biofuel production, providing environmentally benign routes to satisfy the increasing renewable energy demand.

Estimations of sustainable biomass supply available for use in energy applications by 2050 differ widely (40−240 EJ).[2,11–13] Yet, all of them highlight the varied nature of biomass (for example, for a biomass potential of 120 EJ per year, 45 come from forestry, 10 from agriculture, 55 from wastes, and 10 from aquatic sources).[13] Hence, any future conversion technology should have the flexibility to adjust to these multiple and varied feedstocks.

Among the investigated technologies to convert biomass,[14–17] gasification, i.e., the thermochemical conversion by partial oxidation at high temperatures of a solid carbonaceous feedstock to a gaseous product, is the most promising route for biomass valorization as it combines two main

Instituto de Ciencia y Tecnología del Carbono (INCAR), CSIC, Francisco Pintado Fe 26, 33011 Oviedo, Spain. E-mail: victoria.gil@incar.csic.es

Laboratory of Molecular Simulation (LSMO), Institut des Sciences et Ingénierie Chimiques (ISIC), École Polytechnique Fédérale de Lausanne (EPFL), Valais, Rue de l'Industrie 17, Sion CH-1951, Switzerland. E-mail: berend.smit@epfl.ch

The Research Center for Carbon Solutions (RCCS), School of Engineering and Physical Sciences, Heriot-Watt University, EH14 4AS Edinburgh, UK

advantages: high flexibility in terms of feedstock, and versatility to produce different energy carriers. The gaseous product from gasification, syngas, can be used as fuel gas for heat and electricity generation, and as a feedstock for the production of hydrogen, biofuels, and chemicals.[18–20] However, although significant progress has been achieved,[19–26] biomass gasification remains at the development stage as there are still challenges for the wide implementation of biomass gasification plants, which are mainly related to feedstock availability and variability, technology efficiency, and cost-effectiveness compared to other energy sources.[27–31] Among others, the variable chemical composition and properties of biomass is one of the main technical and economic challenges to overcome for the commercial deployment of this technology.[20]

The broad range of possibilities in gasification processes related to the versatility of the gaseous product, *i.e.*, syngas, raises many practical questions such as "What is the most promising use of a particular type of biomass?", or "Should we direct this biomass to a gasification plant that generates energy or power, one that produces methanol, or one that converts the biomass to methane?" Given the diversity of available biomass

feedstocks, identifying the optimal biomass gasification application systematically is essential from a practical point of view. At present, we have to reside to very time-consuming experimental testing[32] to answer this question. We have, at best, some scattered empirical knowledge that can guide us on which application would be optimal for a given biomass.

Indeed, we can find numerous experimental studies on biomass gasification in the literature. At the beginning of 2023, the search query "(biomass OR waste) AND (gasification)" returned over 28 400 records from the Web of Science. However, each of these studies is limited to very specific operating conditions associated with the equipment (and the biomass type) that an experimental group has available. Fig. 1 gives us an idea of the wide range of process conditions used in gasification studies in the literature. For the gasification process, we included the gasification types that can be representative of those that are most studied in research articles: steam gasification, which is carried out at high temperatures utilizing steam as a gasifying agent; supercritical gasification, which refers to the gasification process that occurs in a supercritical fluid state; hydrothermal gasification, a specific type of supercritical gasification that operates at lower temperatures and higher pressures compared to conventional gasification in the presence of supercritical water; and plasma gasification, which is performed at very high temperature achieved by an electric plasma arc. Other gasification processes (such as conventional air gasification, integrated pyrolysis-gasification, or chemical looping gasification) will also be included in the number of records shown in this figure when the search is limited to the type of reactor, gasifying agent, or biomass type. From these works we can, at most, conclude that syngas composition in gasification processes is highly dependent on the type of gasifier, gasifying agent, and operating conditions, as well as the type of biomass. This widespread knowledge makes it extremely difficult to draw conclusive trends that allow us to link a given biomass to a particular application unequivocally.

On the other hand, as we lack a detailed understanding of the reaction mechanism and kinetics due to the complexity of biomass, the conventional theoretical approaches are of little use. This has motivated research groups to explore data-driven approaches.[33–36] Some works use gasification data obtained from thermodynamic simulation studies,[37,38] making easier to create datasets compared to experiments, which can provide a wider overview of the process performance, but the results can vary considerably from real gasifiers. Other studies collect data from several works in the literature and usually focus on the effect of the process operating parameters on the gasification outputs.[34,39–44] We focus on the detailed study of the effect of biomass properties on the gasification process performance, which is still unclear in the literature. From a machine-learning perspective, this is an interesting question as we have to deal with a relatively small dataset. Since biomass conversion studies are very time-consuming, limited data will be a common theme in all of them. In this work, we, therefore, focus on the development of reliable machine-learning models from small datasets.
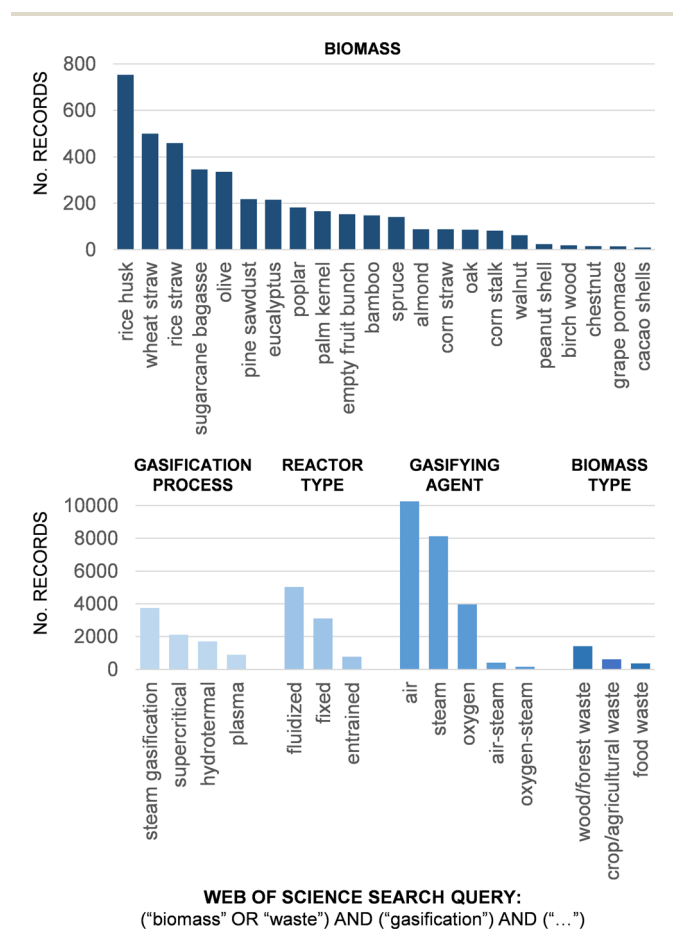

Fig. 1 Diversity of process conditions used in gasification studies in the literature. The *y*-axis shows the number of records returned by the search query "("biomass" OR "waste") AND ("gasification") AND ("…")" from the Web of Science at the beginning of 2023 (search based on title, abstract and indexing).

Machine learning approach

To develop our model, we used data from our bench-scale gasifier published in a previous study[32] and novel data to test our model. Details of the experimental procedure to obtain the data are shown in Section 1 of the ESI.† We studied the gasification of ten types of lignocellulosic biomass from different origins: pine sawdust, chestnut sawdust, torrefied pine sawdust, and torrefied chestnut sawdust (woody materials); almond shells, cacao shells, grape pomace, olive stones, and pine kernel shells (seasonal food industry wastes); and pine cone leaves (forest waste). From the gasification experiments, the gas composition (H_2, CO, CH_4, and CO_2 volume concentrations) and the gas yield (GAS) are experimentally obtained (see Section 1 of the ESI†). We also calculated the total combustible gas concentration (COMBgas) as the sum of the gases with an energy value (H_2, CO, and CH_4). An exploratory analysis of the data used in this work is shown in Section 2 of the ESI.†

We used a multioutput coregionalized Gaussian process regression (GPR) model that is able to capture complex non-linear relationships using only a limited amount of data.[45] In Section 3 of the ESI,† we show the performance of the XGBoost regressor model, used as a baseline (R^2 estimated values are 4 to 23% lower for XGBoost compared to GPR). A dataset of 30 samples was used to develop the model in this study. We used the leave-one-out cross-validation (LOOCV) technique. Therefore, we trained as many models as we have datapoints (N) and then used $N - 1$ points for training and 1 point for testing. More details can be found in Section 4.1 of the ESI.† In contrast to many other machine learning models, the GPR model does not provide us with a simple point estimate but rather a full posterior distribution, providing uncertainty estimates.[46,47] In such a multioutput model, we predict at the same time the most relevant gasification output variables that are typically measured experimentally (combustible gases concentration and gas yield).

Table 1 shows the features and targets used to build the model. As input features of our model, we use parameters that characterize the process and parameters that describe the type of biomass, as they are the main variables that affect the gasification results. The process parameters are gasification temperature (T), steam-to-air (SA) ratio, stoichiometric ratio (SR), and steam-to-biomass ratio (SBR), while biomass is described by its C, H, O, ash, volatile matter (VM), and fixed carbon (FC) contents, its higher heating value (HHVbiom), and its moisture content (MC). Therefore, this work will be limited to the fluidized bed gasifier case, reducing the variability associated to the reactor design and hydrodynamics. The

applicability of the model is by design limited to our setup, and the operating parameter ranges tested in this work are shown in Section 2.2 of the ESI.† As output parameters of our model, we use those variables most typically measured in gasification experiments, such as the volume concentration of H_2, CO, total combustible gas, and also the gas yield (GAS). We can also estimate the volume concentration of CH_4 from these predicted outputs.

Results and discussion

Prediction of gasification outcomes

To test the ability of our machine learning model to predict new data and evaluate its predictive performance, we use so-called leave-one-out cross-validation (LOOCV), a technique that is particularly well-suited for small datasets.[48] Importantly, we perform an additional experimental validation of the model to verify its performance and support the reliability of the model predictions. In Sections 3 and 4 of the ESI† we present a detailed evaluation of the predictive performance of the model. Fig. 2 shows the LOOCV results (blue error bars), plotting the predictions of our model against the actual values for the test points. The coefficient of determination, R^2, shows values between 0.82 and 0.98, indicating a good predictive performance.

Additionally, the results obtained by the model predictions were validated with the experimental results of new gasification experiments that we conducted as part of the present study using two biomasses not previously used for the model training. These results are shown in Fig. 2 with the red error bars, where we can compare the experimental and predicted results for the new experiments. These results show a good agreement between the experimental results and the corresponding values predicted by the model, showing that the model works well on new, never-before-seen, biomasses.
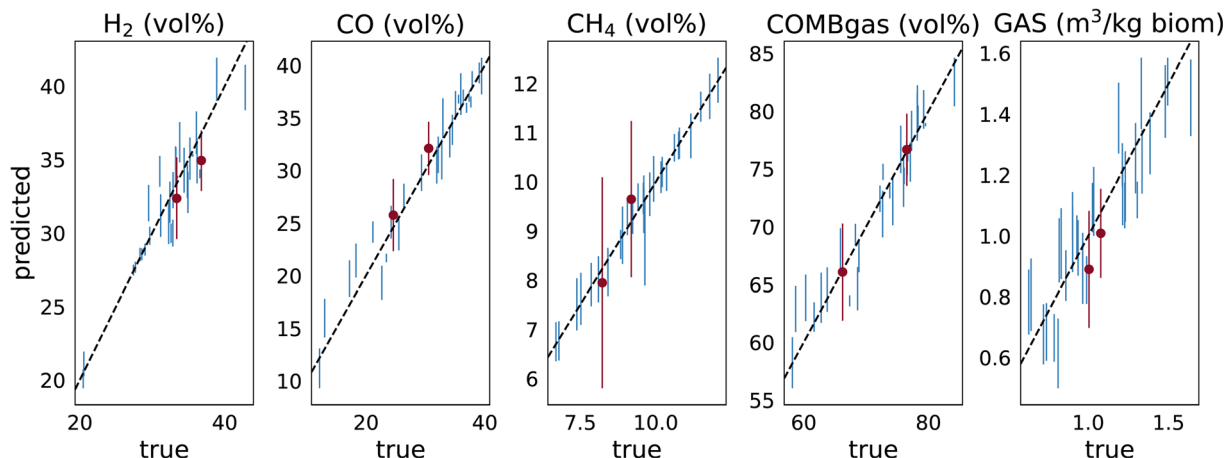
Feature importance analysis

Our machine learning results allow us to carry out a feature importance analysis, which gives us some insights into the process. We computed the SHapley Additive exPlanations (SHAP) feature importance values.[49] The result of this analysis for the most important features is shown in Fig. 3, where the abscissa shows the importance (positive SHAP values indicating higher predictions compared to the average case), and the ordinate shows some of the input features of our model, ordered by overall importance. The color of the points shows the value of those features (red indicates a high feature value).
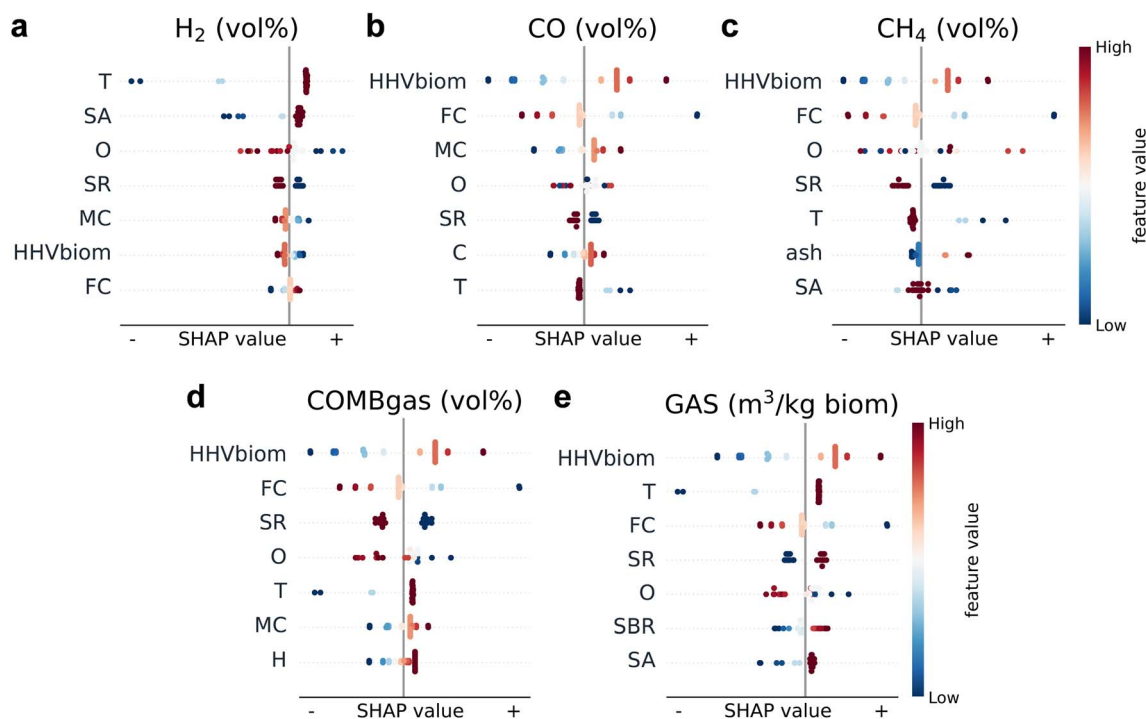
Table 1 Features and targets used to develop the machine learning model

		Variables
Features	Process parameters	T (K), SA, SR, SBR
	Biomass properties	C (%), H (%), O (%), ash (%), VM (%), FC (%), HHVbiom (MJ kg^{-1}), MC (%)
Targets		H_2 (vol%), CO (vol%), CH_4 (vol%), COMBgas (vol%), GAS (m^3 kg^{-1} biom)

Fig. 2 Performance evaluation of the model using both cross-validation and experimental validation. Blue error bars represent the leave-one-out cross-validation (LOOCV) results (test points). Since our dataset is small, we use the leave-one-out strategy for cross-validation, *i.e.*, we train as many models as we have datapoints (N) and then use $N − 1$ points for training and 1 point for testing. Here we show the predictions *vs.* the actual values for the N test points. Error bars show the predicted standard deviations, for CH_4 we compute the error bars considering the error propagation and covariance (see Section 4.5 of the ESI†). R^2 (mean and standard deviation of 15 runs): $H_2 = 0.892 \pm 0.005$, $CO = 0.960 \pm 0.001$, $CH_4 = 0.978 \pm 0.000$, COMBgas $= 0.925 \pm 0.002$, GAS $= 0.819 \pm 0.002$. Red error bars represent the experimental validation results to verify the predictive capacity of the model. For experimental validation, we use new experimental results obtained in our gasifier from the gasification of walnut shells (WS) ($T = 1173$ K, SA $= 2.33$, and SR $= 0.13$) and hazelnut shells (HZS) ($T = 1173$ K, SA $= 2.33$, and SR $= 0.25$) biomasses, which are compared with the values predicted by the model.

Fig. 3 Feature importance analysis for the model predictions using SHapley Additive exPlanations (SHAP) summary plots of the gasification outputs. The figures show on the x-axis the SHAP values that indicate the impact of a feature on the model output compared to a baseline (vertical lines at $x = 0$). The baseline gives the average predicted output values. The y-axis displays the most relevant features that impact the output the most. The width of the spread of the SHAP values on the x-axis indicates the overall importance of the features on the model output. Red points correspond to high feature values, whereas blue points correspond to low feature values. A positive SHAP value means that the specific feature value leads to a higher predicted output value than the baseline prediction. If a feature is irrelevant, all dots, irrespective of color, are on the baseline. For the SHAP analysis, we used all the points in our dataset, which were used to build the coregionalized GPR model. T: temperature (K), SA: steam-to-air ratio, SR: stoichiometric ratio, SBR: steam-to-biomass ratio, C: carbon (wt%), H: hydrogen (wt%), O: oxygen (wt%), FC: fixed carbon (wt%), HHVbiom: biomass higher heating value (MJ kg^{-1}), MC: moisture content (wt%). (a) H_2 predictions. (b) CO predictions. (c) CH_4 predictions are computed by predicting the combustible volume fraction along with CO and H_2. (d) Combustible gas (COMBgas) predictions. (e) Gas yield (GAS) predictions.

Let us first focus on the gasification process parameters. Experimentally, it is well-known that gasification temperature (T) is one of the most relevant variables for the gasification process outcomes.[50,51] It is encouraging that our model correctly indeed identifies T as one of the most important features. The SHAP plot for the H_2 concentration (Fig. 3a) shows that the most important feature for the prediction of this output is precisely the gasification temperature, and that high temperatures have a positive influence on this variable, while very low gasification temperatures (dark blue color points) have a marked negative effect. This can be explained because higher temperatures favor the endothermic water gas and steam reforming reactions according to Le Chatelier's principle, favoring the conversion of formed methane and char produced during the gasification process.[52] A similar effect of the gasification temperature is found for the gas yield (Fig. 3e), also highlighting the negative effect of low gasification temperatures on the gas production.[53] Higher temperatures can favor the production of gas during the biomass devolatilization, as well as promote cracking reactions of secondary hydrocarbons, tars, and char, together with steam reforming and gasification reactions that increase the gas production.[54,55]

In contrast, the gasification temperature has a negative effect on the predicted CO (Fig. 3b) and CH_4 (Fig. 3c) concentrations. At higher temperatures a higher CO production by endothermic reactions can shift the WGS equilibrium towards the consumption of CO.[52,54] Higher temperatures also favor the endothermic steam methane reforming reaction, decreasing the CH_4 content. This agrees with the expected opposite behavior of H_2 and CH_4 during gasification. We can therefore conclude that higher gasification temperatures favor the H_2 production at the expense of CO and CH_4.

The steam-to-air (SA) ratio has also a high importance for the H_2 concentration prediction (Fig. 3a). The feature importance analysis indicates that high values of SA have a positive effect on the H_2 content, while negative importance values are shown by low SA ratios. This confirms the importance of performing the gasification process with steam to obtain a high production of H_2. However, a negative influence of SA on the prediction of the CH_4 concentration is shown, with a slightly lower importance than on the H_2 production. This might be explained because higher steam content favors the reforming reactions, increasing the H_2 production and decreasing the CH_4 production, but, in addition, steam also favors the WGS reaction, increasing H_2, where CH_4 is not involved.[54,55]

From the feature analysis of the process operating conditions, we can conclude that our model correctly captures the experimental trends. We now focus on the biomass properties. The effect of the biomass characteristics on the gasification outputs is studied far less. Our feature importance analysis shows that the most important biomass property is the biomass calorific value (HHVbiom) (Fig. 3). This is an interesting result, as in previous studies in the literature HHVbiom is not often included as a biomass property to study, and only the importance of the biomass C content is usually highlighted.[56–58] The SHAP values show that all outputs, except H_2, increase with the increase in HHVbiom (Fig. 3b–e). In the case of H_2, the biomass calorific value has a relatively lower importance for the prediction of the H_2 concentration, which makes sense as it favors the production of the other gases.

The SHAP plot also shows a relevant negative importance of the fixed carbon (FC) content on the prediction of the CO, CH_4 and combustible gas concentrations, and also of the gas yield. The biomass FC content is closely related to the volatile matter (VM). From this, we can deduce a positive influence of the biomass VM content on these outputs. In contrast, the biomass FC content shows a positive importance on the H_2 concentration prediction. These results indicate that low FC values, *i.e.*, high volatile matter (VM) values, favor the CO and CH_4 production, but not the H_2 production. This can be explained because CO and CH_4 are generated during the devolatilization step of the biomass gasification process.[57] A higher calorific value of the biomass, HHVbiom, could also favor the devolatilization process by increasing the gas phase temperature. Finally, a relevant result from the feature importance analysis is also the negative influence of the biomass O content on the prediction of the H_2 concentration. We can, therefore, deduce that high biomass FC contents and low O contents could be related with a higher H_2 production. On the other hand, the effect of the biomass moisture content is not very relevant in our study since biomasses with significantly high moisture concentrations were not used in this work. In Section 5 of the ESI† we provide a more detailed feature importance analysis, also analyzing partial dependence plots.

Ranking biomasses for different applications

Over the last few years, we have seen in our laboratory an increase in requests from all kinds of producers of biomass, as well as from industries as potential users, to test their biomass in our gasification setup and figure out a potential application. In practice, testing a new type of biomass typically requires a few weeks of work as we need to dry, grind, and sample the biomass to ensure that we have reproducible results. Hence, an important practical application of our model is a simple classification of whether a particular type of biomass is expected to be good for energy production or for conversion into chemicals or fuels. To showcase this application, we use our model to predict the gasification outputs for a number of biomasses whose characteristics we extracted from the literature data.[59–69]

To select the key performance indicator (KPI) parameters that help us to choose the best use for a given biomass, we estimated the molar H_2/CO ratio, the gas calorific value (HHVgas) and the gas energy yield (E_{yield}) as described in the Methods section. The gas energy yield, E_{yield}, is the best key performance indicator (KPI) of the power production from biomass gasification, since it accounts for the gas calorific value and the conversion efficiency (see Section 6 of the ESI† for more details on the selection). Likewise, H_2/CO ratio was chosen as the best KPI for synthesis of fuels/chemicals that require high H_2 concentrations.

In Fig. 4 we represent the predictions of the two selected KPIs for all biomasses gathered from the literature. Here, we can see that some biomasses give high values of E_{yield}, while other
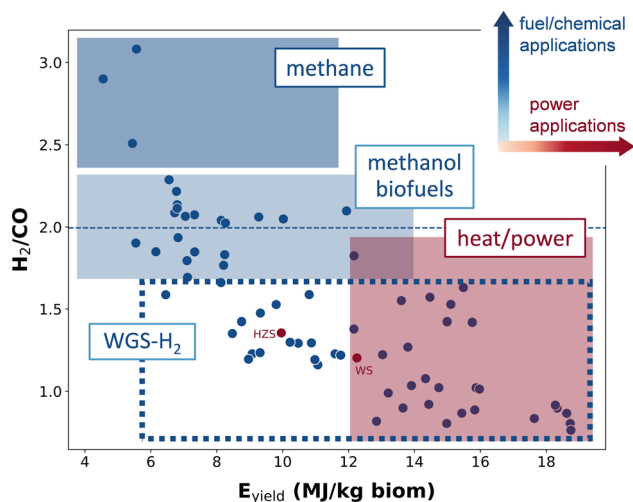
Fig. 4 Selection of biomass applications from the predictions of the key performance indicators (KPIs), E_{yield} and H_2/CO ratio, for all biomasses gathered from the literature. Predictions are estimated under fixed gasification operating conditions of $T = 1173$ K, SA = 2.33, and SR = 0.25. Groups are defined according to the gasification gas requirements for different applications. WGS: water gas shift. Red points represent the model predictions for the new gasification experiments carried out in our gasifier with walnut shells (WS) ($T = 1173$ K, SA = 2.33, and SR = 0.13) and hazelnut shells (HZS) ($T = 1173$ K, SA = 2.33, and SR = 0.25) biomasses.

biomasses produce high values of H_2/CO ratio. This allows us to classify the biomasses into different groups. A high E_{yield} is advantageous to use a biomass for heat or power generation, while a high value of H_2/CO ratio is required to use a particular biomass for synthesis of fuels or chemicals. The synthesis of biofuels by the Fischer–Tropsch process requires a H_2/CO ratio in the syngas around two.[70,71] Likewise, H_2/CO ratio for methanol synthesis should also be set around two (1.7–2.3).[70,72] However, the synthesis of methane by the methanation reaction needs a higher H_2/CO ratio.[70] The stoichiometric H_2/CO ratio is three for the CO methanation reaction, and it has been shown that high H_2/CO ratios improve methanation activity.[73] Based on the potential biomass application, we can classify the studied biomasses in different groups, as shown in Fig. 4: (i) biomasses with high E_{yield}, (ii) biomasses with H_2/CO around 2, (iii) biomasses with $H_2/CO > 2$, and (iv) biomasses with low H_2/CO ratio and low E_{yield}.

These findings pose the question of why some biomasses produce a gasification gas with a higher energy value, while others give a gas with higher hydrogen content. We perform a k-means cluster analysis[74] to find the groups of biomasses that share similar gasification results. Fig. 5a shows the four clusters of similar biomasses. We plot the groups of biomasses as a function of the E_{yield} and H_2/CO ratio. Then, to find relationships between the biomass characteristics and the gasification outputs, we also represent those clusters as a function of the biomass properties in Fig. 5b. Details of the cluster analysis, and the distribution of the different biomass types into the groups, can be found in Section 7 of the ESI.†

The results show that if we want to use biomass for hydrogen-based applications, we need to focus on group 1 (dark

blue) of Fig. 5a, which includes the biomasses that produce the highest H_2/CO ratio. In Fig. 5b, we see that these biomasses are characterized by low contents of carbon, VM and oxygen, relatively low calorific value (HHVbiom), but relatively high FC content. Some biomasses in group 1 are cotton stalks, vine shoots, pineapple waste, cotton seed husks, peanut shells, cacao shells, or sunflower seed shells.

In contrast, group 3 (dark red) in Fig. 5a includes biomasses potentially more convenient for power-based applications. These biomasses give the highest E_{yield}, and they are characterized by high carbon, VM and oxygen contents, relatively high calorific value (HHVbiom), but low FC content (Fig. 5b). Biomass types in group 3 are wood (e.g., pine, beech, poplar, eucalyptus, salix, and willow) and fruit pits (e.g., apricot, prune, and olive).

On the other hand, group 2 (light blue) in Fig. 5a includes biomasses that produce H_2/CO ratio lower than 1.5 and moderate E_{yield}, and they are therefore characterized by intermediate values of the biomass properties between groups 1 and 3 (Fig. 5b). Some of these biomasses are sugarcane bagasse, wheat straw, barley straw, switchgrass, coconut shell, and forest residue wood. For those biomasses that are not particularly promising for either biofuel/chemical or power applications as they produce a gasification gas with low E_{yield} and low H_2/CO ratio (see Fig. 4), we could add a downstream water gas shift (WGS) reactor to our plant to increase the hydrogen content of the gas. The water gas shift reaction converts CO to H_2, increasing the hydrogen concentration. Although an economic evaluation would be required in this case, since both capital and operational costs increase, this could be interesting if it allows us to recycle some available biomass or residual organic material. The use of a WGS reactor to convert the syngas can also be interesting if the production of hydrogen gas is our objective, for example, for the subsequent synthesis of ammonia.

We also have biomasses that produce high E_{yield} and relatively high H_2/CO ratio in group 4 (grey) (Fig. 5a). These are characterized by remarkably high carbon content and calorific value, which favor energy production, but also by high FC and low oxygen contents, which favor hydrogen production (Fig. 5b). Biomasses in group 4 are grape seeds, pine bark, and torrefied woods (beech, eucalyptus, pine, poplar, and spruce). Torrefaction involves the heating of the biomass under an inert or O_2 impoverished atmosphere under mild conditions (200–300 °C) to improve its handling, transportation, and storage characteristics.[75] During torrefaction light volatiles are released, while carbon content and energy density increase, which aligns with the properties of the torrefied biomasses in this study, i.e., high carbon and FC contents (due to the release of volatile matter) and high calorific value, which are the main properties characterizing group 4. These results indicate that our model captures the different gasification behavior of this type of biomass due to its particular characteristics. Our results also show that biomasses that give high H_2/CO ratios (>2) usually produce low E_{yield}, which means a lower conversion efficiency of the feedstock. However, in the case of torrefied woods, the gasification process produces relatively high values for both KPIs. Thus, we could use the torrefaction process to increase the
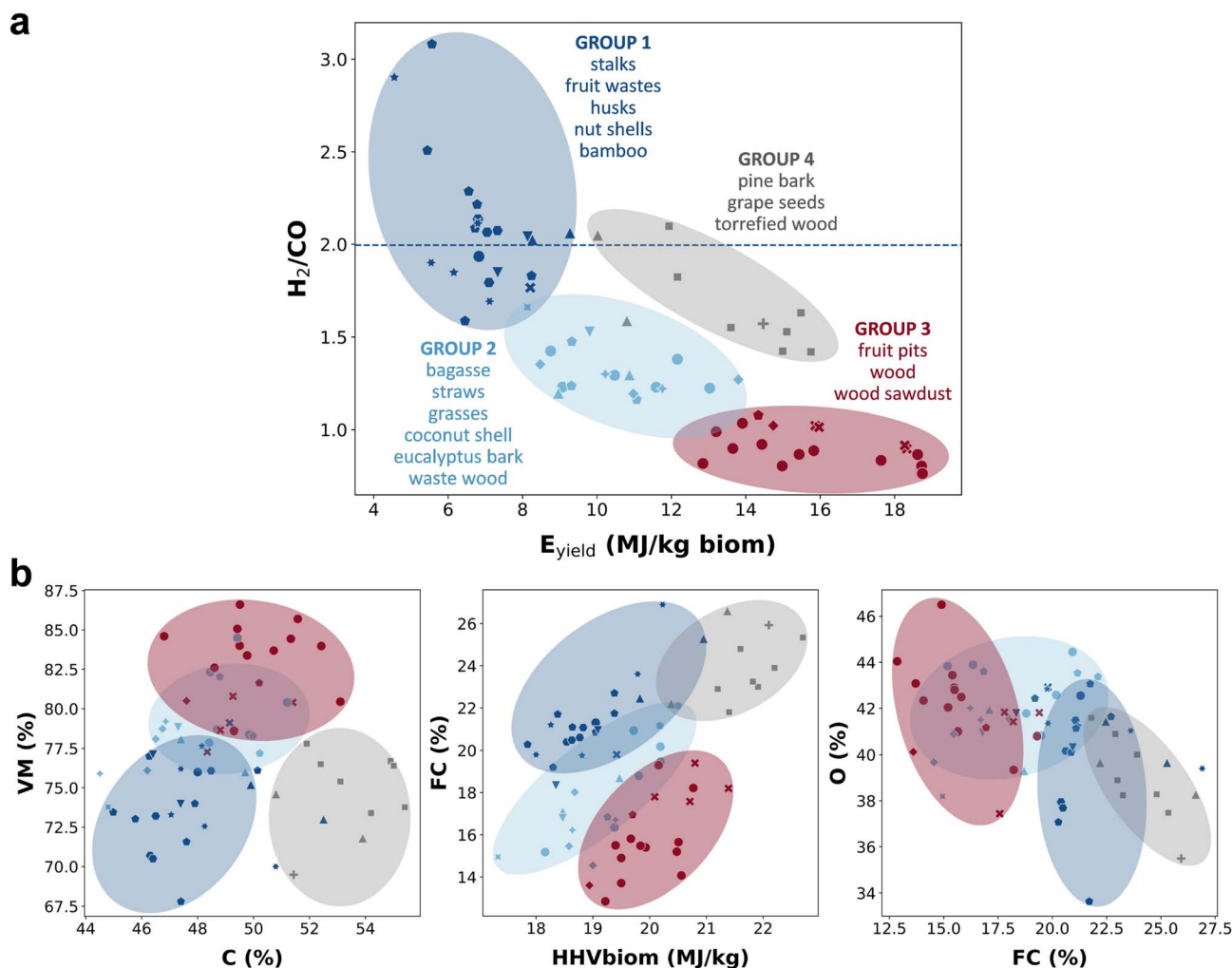
Fig. 5 *K*-Means cluster analysis based on the gasification outputs. Gasification operating conditions: $T = 1173$ K, SA = 2.33, and SR = 0.25. (a) Biomass groups identified by the cluster analysis are plotted as a function of the E_{yield} and H_2/CO ratio values. (b) Biomass groups identified by the cluster analysis are represented as a function of different biomass properties.

hydrogen production for the synthesis of biofuels or chemicals, keeping a satisfactory process conversion. This reveals the torrefaction process as a promising biomass pretreatment before gasification that is worthy of further research.

Optimization of the process operating conditions

Our model can also be used to optimize the operating conditions in the gasifier. In Section 8 of the ESI† we show how the gasification gas characteristics change for different gasification temperatures and steam-to-air ratios, which are the most relevant process variables according to the feature importance analysis. Our model can help us to optimize the gasification conditions for different biomasses.

We show here the case study for biofuel synthesis since it is one of the most attractive biomass conversion routes under research, related to the deep reductions in the carbon emissions from the aviation sector needed by 2050.[76] To carry out the optimization of the process operating conditions, we created a grid of values of the input variables that was fed directly into

the model to obtain the predictions, since we based the optimization on a target H_2/CO ratio of 2 as a function of the values of the gasification temperature and steam-to-air ratio. Fig. 6a shows the value of H_2/CO ratio as a function of the gasification temperature (T) and the steam-to-air (SA) ratio for some biomasses (from groups 1 and 4). All points located between the dashed black lines represent different combinations of T and SA ratio that give a H_2/CO ratio around two (1.9–2.1), which is needed for the synthesis of biofuels by the Fischer–Tropsch process. Under such conditions, these biomasses could be used in a given gasifier to produce biofuels.

Fig. 6b shows how we can also find the specific process conditions for a given biomass that give a gasification gas with a H_2/CO ratio of two. We have estimated the optimum gasification temperature for several biomasses when using a SA ratio of 2.33, and also the optimum SA ratio for each biomass using a gasification temperature of 1173 K. All these biomasses could be gasified in a range of temperatures of 1034–1168 K and SA of 2.33 to obtain a H_2/CO ratio optimal for biofuel synthesis.
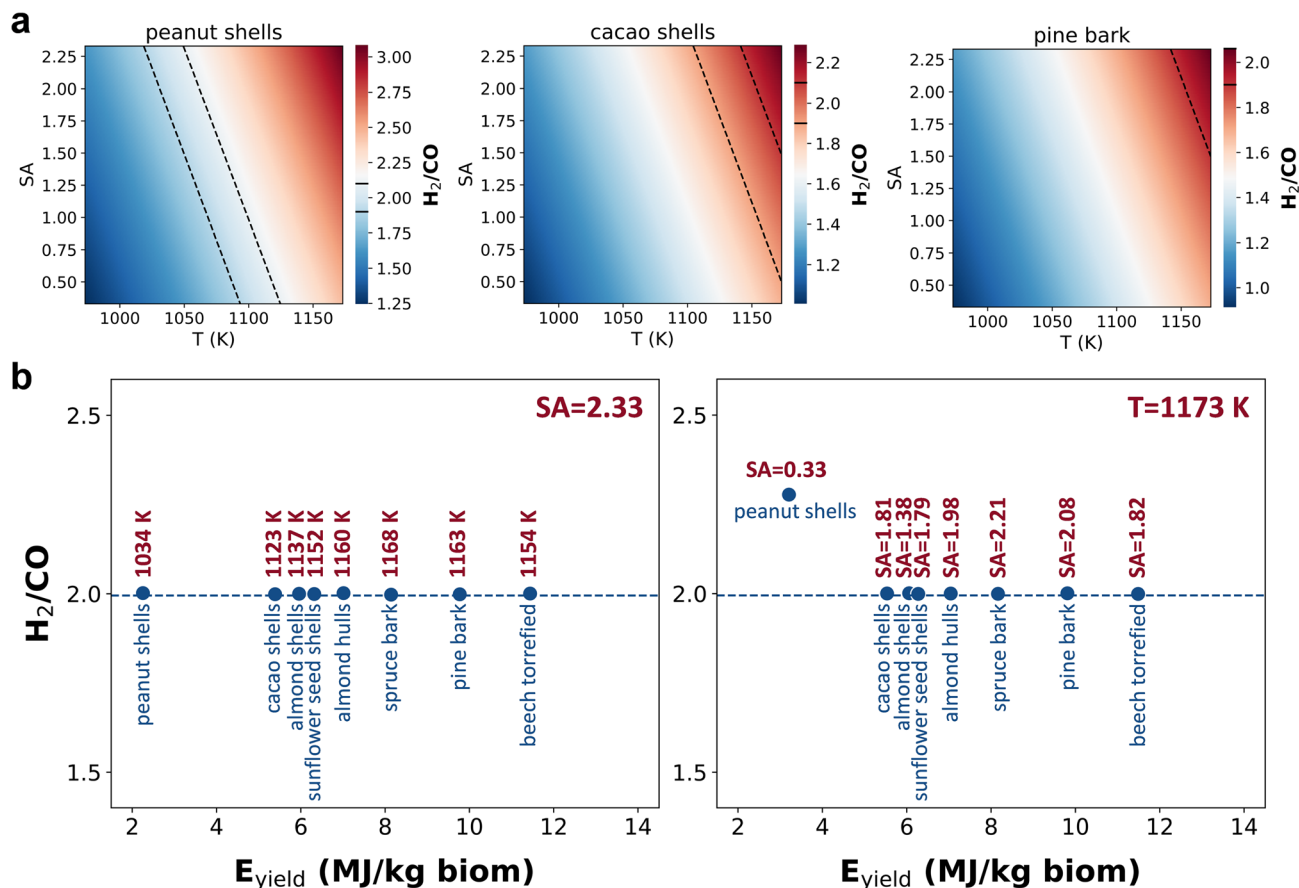
Fig. 6 Optimization of the operating conditions, gasification temperature (T) and steam-to-air (SA) ratio, to produce a gasification gas suitable for biofuel synthesis from different biomasses. (a) Optimum conditions of T and SA that produce a gasification gas with a H_2/CO ratio around two (1.9–2.1) are shown between the dashed black lines for peanut shells, cacao shells and pine bark. (b) Optimum conditions to produce a gasification gas with a H_2/CO ratio of two. The optimum temperature when using a SA ratio of 2.33 has been estimated for different biomasses. The optimum SA ratio when using a gasification T of 1173 K has also been estimated for those biomasses.

Likewise, they could be gasified in a range of SA ratios of 1.38–2.21 at 1173 K. In the case of peanut shells, the H_2/CO ratio is still higher than two even when using the lowest SA ratio studied (0.33), which indicates that the temperature needs to be lower than 1173 K to obtain the selected ratio from this biomass. With this model, we can therefore decide which biomasses could be potentially treated in the same gasifier throughout the year considering our planned final use of the gasification gas for producing biofuels. This will help us to plan the exploitation of a gasifier according to the seasonal biomass availability and/or cost.

Conclusions

Biomass conversion by gasification processes is a typical case of a system that is too complex for conventional mechanistic studies and too time-consuming to obtain a large amount of experimental data. Surprisingly, our work shows that even from relatively small datasets, machine-learning models allow us to establish meaningful correlations between the different potential products from the gasification reaction and the properties of the biomass. That our model correctly predicted the

gasification outcomes for two new biomasses not used previously for the development of the model is an important illustration of the potential of this approach.

From a practical point of view, our model not only correctly predicts the gasification outcomes from different biomasses, but it also matches them with the most promising energy application, for example, heat and power generation or synthesis of biofuels. The model also allows for tailoring the gasification products to the specific requirements of an application, enhancing the flexibility of the process. This will help manage the seasonal biomass availability and a steady feedstock supply. These aspects are crucial for the economic feasibility and successful deployment of gasification plants that are required to fulfill the increasing future demand for renewable and sustainable energy generation.

Likewise, the model developed here opens many avenues for future work as one can expect that many groups have datasets of similar size. For example, the biomass gasification outcomes depend on the characteristics of the reactor, but basic physical and chemical principles are transferable. Thus, collecting similar-size datasets for different types of reactors would allow us to apply techniques such as meta-learning to leverage those

many small datasets and develop a meta-model. Such a model can create good starting points for future optimization campaigns to address the impact of the type of gasifier. This would be an important example of how collective knowledge can be used. In turn, this highlights the importance of promoting open datasets collecting experimental results.

If such a data-driven approach was used systematically in experimental studies, we could contribute to accelerating the development of processes and technologies by providing a more comprehensive understanding of complex processes—through a broader range of insights and predictions—, more informed decision-making, and more efficient scaled-up processes.

Methods

In this section, we provide a short summary of the main methods used in this work. More details can be found in the ESI.†

Features and outputs

Twelve variables were used as input features to build our model, including process operating variables and biomass properties. Four process variables were included: temperature (T), steam-to-air (SA) ratio, stoichiometric ratio (SR), and steam-to-biomass ratio (SBR). Eight characteristics of the biomass were also considered: C, H, and O contents (wt%, dry basis, derived from the ultimate analysis); ash, volatile matter (VM), and fixed carbon (FC) contents (wt%, dry basis, derived from the proximate analysis); higher heating value (HHVbiom); and moisture content (MC) (wt%).

Four outcomes of the gasification process were used as outputs of the model: H_2 concentration (vol%), CO concentration (vol%), combustible gas concentration (vol%) (COMBgas), and gas yield (GAS). COMBgas is the addition of the three main combustible gases obtained from the gasification process (*i.e.*, CO, H_2, and CH_4). The CH_4 concentration (vol%) was estimated from the predicted outputs of the model.

All features and outputs were z-score standardized using the mean and standard deviation of the training set (using the scikit-learn Python package[77]).

In addition, from the experimentally measured gasification outputs, we also estimated other gasification outcomes that can be relevant to evaluate the performance of the gasification process. The molar H_2/CO ratio was calculated from the H_2 and CO concentrations, while the higher heating value (HHV) of the gas obtained (HHVgas),[78] and the energy yield (E_{yield}) were determined as defined by eqn (1), and (2), respectively:

$$\text{HHVgas (MJ Nm}^{-3}) = (11.76 x_{CO} + 11.882 x_{H_2} + 37.024 x_{CH_4}) \times 10/1000 \quad (1)$$

$$E_{yield} \text{ (MJ kg}^{-1} \text{ biom)} = \text{GAS} \cdot \text{HHVgas} \quad (2)$$

where x_i (vol%) represents the volumetric percentage of each component in the dry product gas, and GAS is the gas yield (Nm^3 gas kg^{-1} biom).

Gaussian process regression model

We used the GPy Python library[79] to build and train the Gaussian process regression (GPR) models. We used as the kernel the sum of the Radial basis function (RBF) kernel and the Linear kernel, with automatic relevance determination (ARD) in an intrinsic model of coregionalization (ICM).[80] To evaluate the predictive performance of our model, we used leave-one-out cross-validation (LOOCV), *i.e.*, we train as many models as datapoints (N) we have and then use $N − 1$ points for training and 1 point for testing. More details of the model development and selection can be found in Section 4 of the ESI.† We used the coefficient of determination, R^2, and the root mean squared error, RMSE, to assess the predictive performance of the models. To validate our model, we performed an additional experimental validation with new experimental data from the gasification of biomasses not used previously for the training of the model.

Feature importance analysis

To determine the feature importance we used two different approaches: the SHapley Additive exPlanations (SHAP) technique[49] marginalized over the full dataset to calculate SHAP values, and the partial dependence plots[81] in which the plotted features are marginalized out over the distribution of all features. A more detailed description is given in Section 5.1 of the ESI.†

K-Means cluster analysis

We applied *k*-means clustering[74] to the gasification outputs. This analysis provides an unsupervised grouping of samples with similarity. We find the groups of biomasses that share similar gasification results, and then we look for common biomass properties in each group. *K*-means scikit-learn Python algorithm was used, with *k*-means++ as method for initialization and a number of clusters of 4.[77]

Data availability

The raw data for this study are archived on Zenodo (DOI: **10.5281/zenodo.7358117**). The code for our analysis is available at GitHub (**https://github.com/vgvinter/biomassml**) and archived on Zenodo (DOI: **10.5281/zenodo.7369000**). The version of the code employed for this study is version v0.1.0.

Author contributions

M. V. G. and K. M. J. developed the machine learning approach and wrote the manuscript. M. V. G. carried out the experimental work. All authors contributed to the design of the work, the review and editing of the manuscript, and the discussion of the results.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 IEA, *Net Zero by 2050: A Roadmap for the Global Energy Sector*, International Energy Agency, 2021.

2 IRENA, *Bioenergy for the energy transition: Ensuring sustainability and overcoming barriers*, International Renewable Energy Agency, Abu Dhabi, 2022.

3 G. Luderer, *et al.*, Environmental co-benefits and adverse side-effects of alternative power sector decarbonization strategies, *Nat. Commun.*, 2019, **10**, 1–13.

4 J. C. Solarte-Toro, J. A. González-Aguirre, J. A. Poveda Giraldo and C. A. Cardona Alzate, Thermochemical processing of woody biomass: a review focused on energy-driven applications and catalytic upgrading, *Renewable Sustainable Energy Rev.*, 2021, **136**, 110376.

5 T. Rasheed, M. T. Anwar, N. Ahmad, F. Sher, S. U. D. Khan, A. Ahmad, R. Khan and I. Wazeer, Valorisation and emerging perspective of biomass based waste-to-energy technologies and their socio-environmental impact: A review, *J. Environ. Manage.*, 2021, **287**, 112257.

6 D. S. Bajwa, T. Peterson, N. Sharma, J. Shojaeiarani and S. G. Bajwa, A review of densified solid biomass for energy production, *Renewable Sustainable Energy Rev.*, 2018, **96**, 296–305.

7 M. Ni, D. Y. Leung, M. K. Leung and K. Sumathy, An overview of hydrogen production from biomass, *Fuel Process. Technol.*, 2006, **87**, 461–472.

8 M. Saleem, Possibility of utilizing agriculture biomass as a renewable and sustainable future energy source, *Heliyon*, 2022, **8**, e08905.

9 A. A. Adeleke, P. P. Ikubanni, T. A. Orhadahwe, C. T. Christopher, J. M. Akano, O. O. Agboola, S. O. Adegoke, A. O. Balogun and R. A. Ibikunle, Sustainability of multifaceted usage of biomass: A review, *Heliyon*, 2021, **7**, e08025.

10 S. P. Andersen, B. Allen and G. C. Domingo, *Biomass in the EU Green Deal: Towards consensus on the use of biomass for EU bioenergy. Policy report*, Institute for European Environmental Policy (IEEP), 2021.

11 IEA, *Technology Roadmap - Delivering Sustainable Bioenergy*, International Energy Agency, Paris, 2017.

12 H.-O. Pörtner and *et al.*, *Scientific outcome of the IPBES-IPCC co-sponsored workshop on biodiversity and climate change*, IPBES secretariat, Bonn, Germany, 2021.

13 ETC, *Bioresources within a Net-Zero Emissions Economy: Making a Sustainable Approach Possible*, Energy Transitions Commission, 2021.

14 C. K. Yamakawa, F. Qin and S. I. Mussatto, Advances and opportunities in biomass conversion technologies and biorefineries for the development of a bio-based economy, *Biomass Bioenergy*, 2018, **119**, 54–60.

15 N. M. Kosamia, M. Samavi, K. Piok and S. K. Rakshit, Perspectives for scale up of biorefineries using biochemical conversion pathways: Technology status, techno-economic, and sustainable approaches, *Fuel*, 2022, **324**, 124532.

16 M. Shahbaz, A. AlNouss, I. Ghiat, G. Mckay, H. Mackey, S. Elkhalifa and T. Al-Ansari, A comprehensive review of biomass based thermochemical conversion technologies integrated with CO_2 capture and utilisation within BECCS networks, *Resour., Conserv. Recycl.*, 2021, **173**, 105734.

17 A. Saravanakumar, P. Vijayakumar, A. T. Hoang, E. E. Kwon and W.-H. Chen, Thermochemical conversion of large-size woody biomass for carbon neutrality: Principles, applications, and issues, *Bioresour. Technol.*, 2023, **370**, 128562.

18 Y. Xiao, S. Xu, Y. Song, Y. Shan, C. Wang and G. Wang, Biomass steam gasification for hydrogen-rich gas production in a decoupled dual loop gasification system, *Fuel Process. Technol.*, 2017, **165**, 54–61.

19 Ö. Tezer, N. Karabağ, A. Öngen, C. Ö. Çolpan and A. Ayol, Biomass gasification for sustainable energy production: A review, *Int. J. Hydrogen Energy*, 2022, **11**, 811.

20 S. L. Narnaware and N. Panwar, Biomass gasification for climate change mitigation and policy framework in India: a review, *Bioresour. Technol. Rep.*, 2022, **17**, 100892.

21 S. Valizadeh, H. Hakimian, A. Farooq, B.-H. Jeon, W.-H. Chen, S. H. Lee, S.-C. Jung, M. W. Seo and Y.-K. Park, Valorization of biomass through gasification for green hydrogen generation: a comprehensive review, *Bioresour. Technol.*, 2022, **365**, 128143.

22 S. Mishra and R. K. Upadhyay, Review on biomass gasification: gasifiers, gasifying mediums, and operational parameters, *Mater. Sci. Energy Technol.*, 2021, **4**, 329–340.

23 IEA, *Status report on thermal gasification of biomass and waste 2021*, International Energy Agency, 2022.

24 IEA, *Global database of biomass conversion facilities*, **https://www.ieabioenergy.com/installations/**, accessed June 22, 2023.

25 A. G. Linde, Linde and Forest BtL sign licensing agreement for Carbo-V technology, *Focus on Catalysts*, 2013, 3–4.

26 M. Siedlecki, W. de Jong and A. H. Verkooijen, 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.

27 V. G. Nguyen, T. X. Nguyen-Thi, P. Q. Phong Nguyen, V. D. Tran, Ü. Ağbulut, L. H. Nguyen, D. Balasubramanian, W. A. Tarelko, S. A. Bandh and N. D. Khoa Pham, Recent

advances in hydrogen production from biomass waste with a focus on pyrolysis and gasification, *Int. J. Hydrogen Energy*, 2023, DOI: 10.1016/j.ijhydene.2023.05.049.

28 P. K. Ghodke, A. K. Sharma, A. Jayaseelan and K. P. Gopinath, Hydrogen-rich syngas production from the lignocellulosic biomass by catalytic gasification: a state of art review on advance technologies, economic challenges, and future prospectus, *Fuel*, 2023, **342**, 127800.

29 A. Akbarian, A. Andooz, E. Kowsari, S. Ramakrishna, S. Asgari and Z. A. Cheshmeh, Challenges and opportunities of lignocellulosic biomass gasification in the path of circular bioeconomy, *Bioresour. Technol.*, 2022, **362**, 127774.

30 IEA, *Outlook for biogas and biomethane: Prospects for organic growth*, International Energy Agency, 2020.

31 S. K. Sansaniwal, K. Pal, M. A. Rosen and S. K. Tyagi, Recent advances in the development of biomass gasification technology: a comprehensive review, *Renewable Sustainable Energy Rev.*, 2017, **72**, 363–384.

32 M. González-Vázquez, R. García, M. Gil, C. Pevida and F. Rubiera, Comparison of the gasification performance of multiple biomass types in a bubbling fluidized bed, *Energy Convers. Manage.*, 2018, **176**, 309–323.

33 S. Ascher, I. Watson and S. You, Machine learning methods for modelling the gasification and pyrolysis of biomass and waste, *Renewable Sustainable Energy Rev.*, 2022, **155**, 111902.

34 S. Ascher, W. Sloan, I. Watson and S. You, A comprehensive artificial neural network model for gasification process prediction, *Appl. Energy*, 2022, **320**, 119289.

35 G. C. Umenweke, I. C. Afolabi, E. I. Epelle and J. A. Okolie, Machine learning methods for modeling conventional and hydrothermal gasification of waste biomass: a review, *Bioresour. Technol. Rep.*, 2022, **17**, 100976.

36 M. khan, S. R. Naqvi, Z. Ullah, S. A. A. Taqvi, M. N. A. Khan, W. Farooq, M. T. Mehran, D. Juchelková and L. Štěpanec, Applications of machine learning in thermochemical conversion of biomass-a review, *Fuel*, 2023, **332**, 126055.

37 S. Safarian, S. M. Ebrahimi Saryazdi, R. Unnthorsson and C. Richter, Artificial neural network integrated with thermodynamic equilibrium modeling of downdraft biomass gasification-power production plant, *Energy*, 2020, **213**, 118800.

38 S. Sezer, F. Kartal and U. Özveren, Prediction of chemical exergy of syngas from downdraft gasifier by means of machine learning, *Therm. Sci. Eng. Prog.*, 2021, **26**, 101031.

39 S. Ren, S. Wu and Q. Weng, Physics-informed machine learning methods for biomass gasification modeling by considering monotonic relationships, *Bioresour. Technol.*, 2023, **369**, 128472.

40 J. Y. Kim, D. Kim, Z. J. Li, C. Dariva, Y. Cao and N. Ellis, Predicting and optimizing syngas production from fluidized bed biomass gasifiers: a machine learning approach, *Energy*, 2023, **263**, 125900.

41 J. Li, L. Pan, M. Suvarna and X. Wang, Machine learning aided supercritical water gasification for H_2-rich syngas production with process optimization and catalyst screening, *Chem. Eng. J.*, 2021, **426**, 131285.

42 P. V. Gopirajan, K. P. Gopinath, G. Sivaranjani and J. Arun, Optimization of hydrothermal gasification process through machine learning approach: experimental conditions, product yield and pollution, *J. Cleaner Prod.*, 2021, **306**, 127302.

43 M. Puig-Arnavat, J. A. Hernández, J. C. Bruno and A. Coronas, Artificial neural network models for biomass gasification in fluidized bed gasifiers, *Biomass Bioenergy*, 2013, **49**, 279–289.

44 S. Ascher, X. Wang, I. Watson, W. Sloan and S. You, Interpretable machine learning to model biomass and waste gasification, *Bioresour. Technol.*, 2022, **364**, 128062.

45 K. M. Jablonka, G. M. Jothiappan, S. Wang, B. Smit and B. Yoo, Bias free multiobjective active learning for materials design and discovery, *Nat. Commun.*, 2021, **12**, 2312.

46 M. A. Álvarez, L. Rosasco and N. D. Lawrence, Kernels for vector-valued functions: a review, *Found. Trends Mach. Learn.*, 2011, **4**, 195–266.

47 M. van der Wilk, V. Dutordoir, S. John, A. Artemev, V. Adam and J. Hensman, *A Framework for Interdomain and Multioutput Gaussian Processes*, 2020, http://arxiv.org/abs/2003.01115.

48 K. M. Jablonka, D. Ongari, S. M. Moosavi and B. Smit, Big-Data Science in Porous Materials: Materials Genomics and Machine Learning, *Chem. Rev.*, 2020, **120**, 8066–8129.

49 S. M. Lundberg and S.-I. Lee, in *Advances in Neural Information Processing Systems 30*, ed. Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S. and Garnett R., Curran Associates, Inc., 2017, pp. 4765–4774.

50 A. A. Ahmad, N. A. Zawawi, F. H. Kasim, A. Inayat and A. Khasri, Assessing the gasification performance of biomass: a review on biomass gasification process conditions, optimization and economic evaluation, *Renewable Sustainable Energy Rev.*, 2016, **53**, 1333–1347.

51 H. Song, G. Yang, P. Xue, Y. Li, J. Zou, S. Wang, H. Yang and H. Chen, Recent development of biomass gasification for H_2 rich gas production, *Appl. Energy Combust. Sci.*, 2022, **10**, 100059.

52 X. Ku, H. Jin and J. Lin, Comparison of gasification performances between raw and torrefied biomasses in an air-blown fluidized-bed gasifier, *Chem. Eng. Sci.*, 2017, **168**, 235–249.

53 M. Campoy, A. Gómez-Barea, A. L. Villanueva and P. Ollero, Air-Steam Gasification of Biomass in a Fluidized Bed under Simulated Autothermal and Adiabatic Conditions, *Ind. Eng. Chem. Res.*, 2008, **47**, 5957–5965.

54 J. Gil, M. P. Aznar, M. A. Caballero, E. Francés and J. Corella, Biomass Gasification in Fluidized Bed at Pilot Scale with Steam-Oxygen Mixtures. Product Distribution for Very Different Operating Conditions, *Energy Fuels*, 1997, **11**, 1109–1118.

55 S. Fremaux, S.-M. Beheshti, H. Ghassemi and R. Shahsavan-Markadeh, An experimental study on hydrogen-rich gas production via steam gasification of biomass in a research-scale fluidized bed, *Energy Convers. Manage.*, 2015, **91**, 427–432.

56 M. Dellavedova, M. Derudi, R. Biesuz, A. Lunghi and R. Rota, On the gasification of biomass: data analysis and regressions, *Process Saf. Environ. Prot.*, 2012, **90**, 246–254.

57 G. Mirmoshtaghi, J. Skvaril, P. E. Campana, H. Li, E. Thorin and E. Dahlquist, The influence of different parameters on biomass gasification in circulating fluidized bed gasifiers, *Energy Convers. Manage.*, 2016, **126**, 110–123.

58 I. L. Motta, A. N. Marchesan, R. Maciel Filho and M. R. Wolf Maciel, Correlating biomass properties, gasification performance, and syngas applications of Brazilian feedstocks *via* simulation and multivariate analysis, *Ind. Crops Prod.*, 2022, **181**, 114808.

59 TNO, *Phyllis2, database for (treated) biomass, algae, feedstocks for biogas production and biochar*, https://phyllis.nl/, accessed December 16, 2022.

60 S. A. Channiwala and P. P. Parikh, A unified correlation for estimating HHV of solid, liquid and gaseous fuels, *Fuel*, 2002, **81**, 1051–1063.

61 M. Gil, P. Oulego, M. Casal, C. Pevida, J. Pis and F. Rubiera, Mechanical durability and combustion characteristics of pellets from biomass blends, *Bioresour. Technol.*, 2010, **101**, 8859–8867.

62 A. Veses, O. Sanahuja-Parejo, M. V. Navarro, J. M. López, R. Murillo, M. S. Callén and T. García, From laboratory scale to pilot plant: evaluation of the catalytic co-pyrolysis of grape seeds and polystyrene wastes with CaO, *Catal. Today*, 2021, **379**, 87–95.

63 A. A. Tortosa Masiá, B. J. Buhre, R. P. Gupta and T. F. Wall, Characterising ash of biomass and waste, *Fuel Process. Technol.*, 2007, **88**, 1071–1081.

64 A. Shakya and T. Agarwal, Removal of Cr(VI) from water using pineapple peel derived biochars: Adsorption potential and re-usability assessment, *J. Mol. Liq.*, 2019, **293**, 111497.

65 A. Demirbas, Combustion characteristics of different biomass fuels, *Prog. Energy Combust. Sci.*, 2004, **30**, 219–230.

66 T. Järvinen and D. Agar, Experimentally determined storage and handling properties of fuel pellets made from torrefied whole-tree pine chips, logging residues and beech stem wood, *Fuel*, 2014, **129**, 330–339.

67 P. C. Kuo, B. Illathukandy, W. Wu and J. S. Chang, Plasma gasification performances of various raw and torrefied biomass materials using different gasifying agents, *Bioresour. Technol.*, 2020, **314**, 123740.

68 B. Arias, C. Pevida, J. Fermoso, M. G. Plaza, F. Rubiera and J. J. Pis, Influence of torrefaction on the grindability and reactivity of woody biomass, *Fuel Process. Technol.*, 2008, **89**, 169–175.

69 J. G. Pohlmann, E. Osório, A. C. Vilela, M. A. Diez and A. G. Borrego, Integrating physicochemical information to follow the transformations of biomass upon torrefaction and low-temperature carbonization, *Fuel*, 2014, **131**, 17–27.

70 I. Wender, Reactions of synthesis gas, *Fuel Process. Technol.*, 1996, **48**, 189–297.

71 W. L. Luyben, Control of parallel dry methane and steam methane reforming processes for Fischer-Tropsch syngas, *J. Process Control*, 2016, **39**, 77–87.

72 M. H. Khademi, A. Alipour-Dehkordi and M. Tabesh, Optimal design of methane tri-reforming reactor to produce proper syngas for Fischer-Tropsch and methanol synthesis processes: A comparative analysis between different side-feeding strategies, *Int. J. Hydrogen Energy*, 2021, **46**, 14441–14454.

73 I. Hussain, A. A. Jalil, C. R. Mamat, T. J. Siang, A. F. Rahman, M. S. Azami and R. H. Adnan, New insights on the effect of the H$_2$/CO ratio for enhancement of CO methanation over metal-free fibrous silica ZSM-5: Thermodynamic and mechanistic studies, *Energy Convers. Manage.*, 2019, **199**, 112056.

74 S. P. Lloyd, Least Squares Quantization in PCM, *IEEE Trans. Inf. Theory*, 1982, **28**, 129–137.

75 M. Gil, R. García, C. Pevida and F. Rubiera, Grindability and combustion behavior of coal and torrefied biomass blends, *Bioresour. Technol.*, 2015, **191**, 205–212.

76 A. Karim, M. A. Islam, A. Nayeem and A. Yousuf, in *Sustainable Alternatives for Aviation Fuels*, ed. Yousuf A. and Gonzalez-Fernandez C., Elsevier, 2022, pp 1–25.

77 F. Pedregosa, *et al.*, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

78 AENOR, *UNE-EN ISO 6976: Gas natural. Cálculo del poder calorífico, densidad, densidad relativa e índice de Wobbe a partir de la composición*, 2005.

79 GPy, *GPy: A Gaussian process framework in python*, 2012, http://github.com/SheffieldML/GPy.

80 M. A. Álvarez, L. Rosasco and N. D. Lawrence, *Kernels for Vector-Valued Functions: A Review*, 2011, https://arxiv.org/abs/1106.6251.

81 J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann. Math. Stat.*, 2001, **29**, 1189–1232.