



Interpretable machine learning to model biomass and waste gasification

Simon Ascher^a, Xiaonan Wang^b, Ian Watson^a, William Sloan^a, Siming You^{a,*}

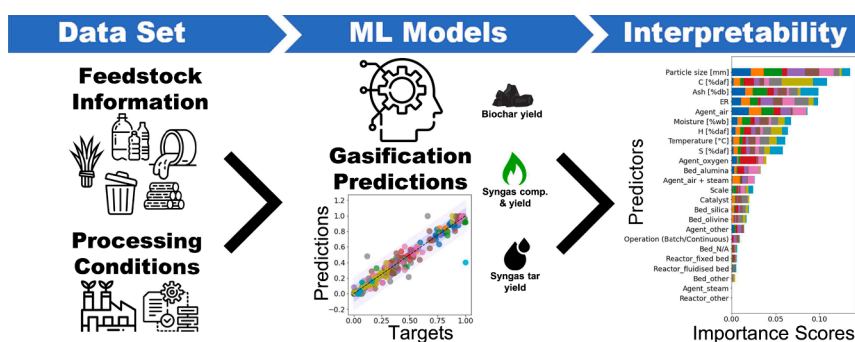
^a School of Engineering, University of Glasgow, University Avenue, Glasgow G12 8QQ, United Kingdom

^b Department of Chemical Engineering, Tsinghua University, Beijing 100084, China

HIGHLIGHTS

- Developed models showed excellent prediction accuracy across 10 key outputs.
- Gradient boosting regression outperformed other model types.
- Feedstock particle size and gasifying agent most strongly influenced predictions.
- Global and local model interpretability methods have been combined.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Bioenergy
Waste-to-energy
Gradient boosting
SHAP (SHapley Additive exPlanations)

ABSTRACT

Machine learning has been regarded as a promising method to better model thermochemical processes such as gasification. However, their black box nature can limit how much one can trust and learn from the developed models. Here seven different machine learning methods have been adopted to model the gasification of biomass and waste across a wide range of operating conditions. Gradient boosting regression has been found to outperform the other model types with a coefficient of determination (R^2) of 0.90 when averaged across ten key gasification outputs. Global and local model interpretability methods have been used to illuminate the developed black box models. The studied models were most strongly influenced by the feedstock's particle size and the type of gasifying agent employed. By combining global and local interpretability methods, the understanding of black box models has been improved. This allows policy makers and investors to make more educated decisions about gasification process design.

1. Introduction

Bioenergy will likely play a critical role in achieving the sustainable development goals set out by the 2015 Paris Agreement. To date, only a small fraction of its potential has been tapped (Lacrosse et al., 2021). Thermochemical methods, such as gasification, are key technologies

highlighted as part of the UK government's Net Zero Strategy. Amongst other reasons, gasification stands out to be an attractive renewable energy technology, as it has, unlike wind and solar, the potential to provide a stable baseload. When combined with carbon capture and storage (i.e. bioenergy with carbon capture and storage (BECCS)), gasification could become a negative emission technology that is highly demanded to fulfil

* Corresponding author.

E-mail address: siming.you@glasgow.ac.uk (S. You).

<https://doi.org/10.1016/j.biortech.2022.128062>

Received 17 August 2022; Received in revised form 27 September 2022; Accepted 29 September 2022

Available online 3 October 2022

0960-8524/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

the 1.5 °C climate goal. Additionally, gasification can aid with reducing the emissions of the hardest-to-decarbonise sectors, such as transport and heating (BEIS, 2021).

Whilst the potential of gasification is clear, its widespread commercialisation and industrial scale implementation is still subject to limited process performance and efficiency (BEIS, 2021). It becomes an important topic to search for new methods and tools towards improved process design. Machine learning (ML) represents one such method. ML is a type of artificial intelligence, which uses a data-driven approach to develop methods which “learn” from provided data. This way models can be developed without any predetermined equations. For this work, ML is a promising method to achieve high accuracy gasification production predictions, which can simplify model-based process design and optimisation. However, existing models often focused on a narrow range of feedstock types and gasifier setups (Ascher et al., 2022b). For instance, Kardani et al. compared five soft computing methods, but focused on municipal solid waste (MSW) gasification only (Kardani et al., 2021).

In addition to low data availability and the narrow scope of existing models, limited interpretability has been another major problem. When first developed, many ML methods have been considered a black box, but more recent efforts to illuminate the inner workings of these methods have led to improved process understanding and analysis (Molnar, 2019). Indeed, interpretability is crucial for allowing researchers to guard against bias in their models and extract maximum knowledge from their data and models, improving the efficiency of data and model utilisation. A variety of methods such as correlation coefficients, hypothesis tests, and variance-based methods have been developed to assess the importance of a model’s variables to aid with interpretability (Wei et al., 2015). A model is considered interpretable when the reasoning behind its predictions can be readily understood by humans.

In the field of bioenergy, conventionally, model-specific techniques predominantly aimed at explaining a model’s global behaviour have been used. Global model interpretability is concerned with a holistic view of the model and how the distribution of model outcomes is affected by the model’s features. For instance, Serrano and Castelló used Garson’s algorithm to explain the variable importance of a tar prediction model for a bubbling fluidised bed gasifier (Serrano and Castelló, 2020). Zhu et al. used Gini importance based variable importance assessment to analyse a random forest (RF) model for the prediction of biochar yield from biomass pyrolysis (Zhu et al., 2019). Zhang et al. used Pearson correlation coefficients, Gini feature importance assessment, and partial dependence plots (PDPs) to study the effect of biomass composition and pyrolysis conditions on bio-oil characteristics (Zhang et al., 2022). The developed RF was found to rely more heavily on the biomass composition than the pyrolysis conditions to make its predictions.

Model-agnostic methods, which aim to separate the explanations from the ML model, have been a more recent development (Lundberg and Lee, 2017; Ribeiro et al., 2016). The key advantage of model-agnostic techniques lies in their flexibility. This way a researcher is free to use any model type and different model types can be compared by using the same interpretability methods. SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) is one such method which has recently been used (Li et al., 2021a; Li et al., 2020). SHAP can be used as a global method to explain the average effect of features on an output, but also as a local method to explain individual predictions. Whilst, in this field, these local methods have not been explored yet, they can powerfully communicate to stakeholders how a model has made a particular prediction. The merits of the interpretability methods selected for this work are discussed in more detail in Section 2.4.

In this work a range of ML methods to model biomass and waste gasification have been developed and systematically compared. Instead of focusing on a narrow range of feedstocks and operating conditions like many existing studies (Ascher et al., 2022b), an expansive data set including an array of different feedstock and gasifier types has been

used. Model performance has been optimised by identifying the best model preprocessing steps for each model type and tuning their hyperparameters using a search algorithm.

As compared to existing interpretability-related studies, the consideration of three different feature importance assessment methods and the use of global and local methods allowed for the creation of a more complete picture of how the developed models made their predictions. Gini, permutation, and SHAP feature importance assessment have been compared, which mitigates limitations of individual methods and reduces uncertainty in the results (Gevrey et al., 2003). The use of local explainability methods to guide the decisions of stakeholders and investors by providing an intuitive way to communicate a ML model’s prediction process has been illustrated. By combining local with global methods, thermochemical process prediction black box models can be better understood. The developed models can predict syngas yield, syngas lower heating value (LHV), and syngas tar content, as well as the char yield and syngas composition in terms of N₂, H₂, CO, CO₂, CH₄, and C₂H_n. These models and their predictions can simplify further gasification system design and analysis such as life cycle sustainability assessment (LCSA) during which a system’s environmental, economic, and social impacts are studied.

2. Materials and methods

2.1. Data collection, preliminary analysis, and predictor selection

The data set used for model development has been presented in greater detail in one of our previous works (Ascher et al., 2022a). It contains information on the feedstock’s composition and its preparation, as well as data describing the gasification system and how it was operated. In total, up to 312 samples were collected from the current gasification literature for a range of different gasifier types, operating conditions, and feedstocks.

Preliminary analysis followed a similar procedure to the one described in one of our previous works (Ascher et al., 2022a). In short, Pearson’s and Spearman’s correlation coefficients were calculated to measure the linear/monotonic relationship between predictors. It is important to identify highly correlated features before model training, as retaining them can limit a model’s performance and hinder feature importance assessment (Li et al., 2021a).

Predictor variables were excluded for several reasons. The feedstock’s N and O content and LHV were excluded due to their high correlation with other predictors. Volatile matter (VM) and fixed carbon (FC) were removed as literature considers them dependent variables (Baruah et al., 2017). Feedstock type and shape were deemed redundant as the same information was already captured by the feedstock’s ultimate and proximate composition and particle size. Finally, several predictors were dropped as they contained too many missing samples. Namely, the feedstock’s cellulose, hemicellulose, and lignin contents, as well as the operating pressure, residence time, and steam to biomass ratio were dropped for this reason.

The following predictors were ultimately used for model development: feedstock C, H, S, ash, and moisture content, feedstock particle size, gasifier temperature, gasifier operation mode (batch/continuous), gasifier scale, equivalence ratio (ER), catalyst usage, gasifying agent, reactor type, and bed material.

All analysis was done in the Python programming language (version 3.8.13).

2.2. Data preparation

Careful data preparation is an essential step before model training. It ensures maximum performance and trust in the developed models. Many ML algorithms require data to be in a uniform and specific format. For instance, missing or invalid entries cannot be processed by most algorithms and thus need to be treated beforehand. In this work, all data

was cleaned, and categorical predictors were encoded using ordinal or one-hot encoding. Multilayer perceptron neural networks and support vector machines (SVMs) prefer features which are normally distributed. Thus, each feature was standardised by removing the mean and scaling samples to unit variance.

Regarding categorical variables, the following variables were ordinally encoded: operation mode (batch or continuous), system scale (lab or pilot), and catalyst use (present or not present). Gasifying agent, bed material, and reactor type were one-hot encoded. Both encoding techniques have been discussed in more detail in (Ascher et al., 2022a).

To study the effects of various data preparation and preprocessing options a range of scenarios was defined. For this, predictor and target data were treated independently. Predictor data was prepared based on one of the following four procedures: (i) Continuous predictors only and drop row if missing value present (CONT + DROP); (ii) Continuous predictors only and mean impute missing values (CONT + MEAN); (iii) Add categorical predictors and drop row if missing value present (ALL + DROP), (iv) Add categorical predictors and mean impute missing values (ALL + MEAN).

The aim of comparing these four options was twofold. Firstly, the effect of adding categorical predictors to the training data set was studied by comparing CONT + DROP and CONT + MEAN to ALL + DROP and ALL + MEAN. Secondly, the effect of mean imputing missing values was studied. This is one of the simplest and most employed methods to deal with missing values. One main benefit of mean imputing missing values, compared to dropping the row is that it does not lead to a reduction in the size of the data set used for model development. During mean imputation, all instances where a sample is missing for a variable are filled with the mean of all samples that are present for that variable.

Target data was only processed for the training set. If missing values were present in the test set a prediction was not made. This way test data remained independent and was not affected by any data preprocessing. The training data was prepared based on one of the following three procedures: (i) Drop row if missing value present (TAR-DROP); (ii) Mean impute missing values (TAR-MEAN); (iii) Impute missing values using RF submodel (TAR-RF).

Methods TAR-DROP and TAR-MEAN follow similar procedures as the methods described for the preparation of the predictor data. Method TAR-RF employs RF submodels to fill missing values. One RF submodel was fitted for each target variable with missing values present. Each submodel was trained using the complete data set and then used to make predictions for the missing target values. Using this method, a size reduction in the data set can be avoided (similar to the mean imputation method) whilst filling the missing instances with more meaningful values than simply the variable's mean. RF was chosen for the submodels due to its strong out of the box performance (Kégl, 2013). A visual explanation of the employed methods is shown in the [supplementary materials](#).

2.3. Model development

2.3.1. Model types

A range of different model types have been compared in this work. One important family of models is tree-based models. On a fundamental level, tree-based models are algorithms which infer simple decision rules from the feature space to predict a target variable (Hastie et al., 2009). In practice, many models employ a collection of trees to reduce the variance associated with using a singular decision tree. RF is one such method introduced by Breiman (Breiman, 2001). It is a bootstrap aggregating, also called bagging, algorithm which averages the results of many independent learners. By averaging the results of many decision trees, the noise problems of individual trees can be circumvented (Hastie et al., 2009). Whilst bagging algorithms average the results of many independent models, boosting algorithms iteratively train and adjust the weights of many weak learners to create a powerful ensemble model. In

this work three different algorithms falling under this category were considered, namely gradient boosting for regression (GBR), XGBoost, and AdaBoost.

SVMs are another class of well-performing algorithms and were also considered in this work. Whilst modern deep neural networks often outperform SVMs or some of the other mentioned model types when vast amounts of data are available, for the size of data set available for this work, the presented algorithms could be among the best performing algorithms in terms of prediction accuracy (Brunton and Kutz, 2017). For this reason, instead of a deep neural network, a simpler multilayer perceptron neural network (ANN) was considered as another alternative. It is worth noting that artificial neural networks have been the most popular ML method to model biomass and waste gasification (Ascher et al., 2022b). The final tested model type is the super learner (SL) concept which combines all previously developed models into one ensemble model (Van Der Laan et al., 2007).

2.3.2. Model optimisation and comparison

Whilst some algorithms, such as RF and AdaBoost, are well known for their good out of the box performance, other algorithms, such as ANN and SVM, require careful hyperparameter tuning to maximise their performance (Kégl, 2013). In the past, trial and error has often been manually implemented to tune ML models; however, more advanced methods have become available to find the best hyperparameter combinations (Ascher et al., 2022b). In this work, a search algorithm has been used to optimise each type of model. Specifically, a parameter grid was defined for each model type. These grids contained the hyperparameters which were to be optimised and possible options/values each parameter could take. The algorithm then searched the parameter space for the best performing combination of hyperparameters. Model performance was assessed by determining the 5-fold cross-validated coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i^o - y_i^p)^2}{\sum_{i=1}^N (y_i^o - \bar{y}^o)^2} \quad (2)$$

where y_i^o and y_i^p are the observed and predicted values. The mean of all predicted values is represented by \bar{y}^o and N is the number of samples. The highest R^2 score indicates a model's best performing hyperparameter combination. The root mean square error (RMSE) is another performance measure given by.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^o - y_i^p)^2} \quad (3)$$

where the variables have the same meaning as the variables used in Eq. (2). It measures the mean difference between observed values and an estimator's predictions.

R^2 and RMSE were used in conjunction as performance measures to quantitatively compare all the model types. They were calculated for test sets and the cross-validated models. For the model development, the data set was split into a training and test set using an 85 % to 15 % split. Another approach for accurately judging a model's generalisation capability is using cross-validation. This work employed k -fold cross-validation, during which the dataset was split into k parts. The model was then trained on $k-1$ folds and tested on the last fold. This process was then repeated k times, so that each fold has been used for testing once. Cross-validation allows one to obtain an independent measure of model performance, whilst using all available data for model development. Here 5-fold cross-validation has been used.

The overall methodology and workflow of this study is illustrated by Fig. 1, where the left-hand side of the figure focuses on the ML model development described in this section.

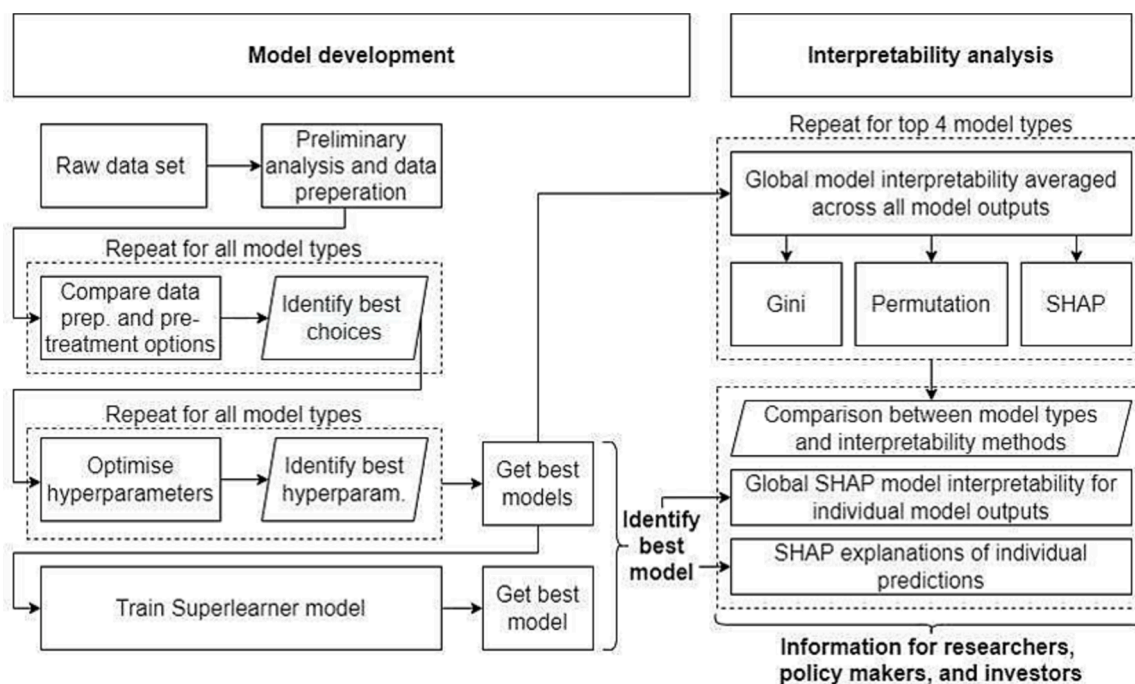


Fig. 1. Flowchart of the methodology and workflow of this study, illustrating the two stages model development and interpretability analysis.

2.4. Interpretability analysis

Model interpretability and explainability are concerned with extracting knowledge about relationships learned by a model or the patterns of the underlying data. Interpretability may be understood as a human's ability to understand the cause of a ML model's prediction. A highly interpretable model is easy to comprehend, and the model's results can be consistently predicted by a human. In contrast, humans struggle to understand the predictions of a model with low interpretability (Molnar, 2019).

Many ML models have been considered a black box, as their inner workings are often poorly understood. One approach to make ML more interpretable is the use of models such as linear regression and decision trees which are inherently interpretable (Molnar, 2019). However, these simple models generally do not offer the same prediction performance as more complex models (Tang et al., 2020). A more recent approach are model-agnostic methods which aim to make any black box model interpretable. One intuitive approach is called permutation feature importance during which a feature's values are randomly shuffled (Breiman, 2001). The resulting increase in the prediction error yields a measure of the feature's importance. This provides a global insight into the model's behaviour and automatically accounts for interaction effects of features (Molnar, 2019).

Another global method uses the Gini index, also known as Gini impurity, which is used by tree-based models to determine where nodes should be split. This provides a straightforward measure to interpret tree-based methods such as RF and GBR. One disadvantage of Gini-based variable importance assessment is that it is biased towards inputs with more categories. This is less of an issue for mostly continuous and largely uncorrelated features. Furthermore, this method is not applicable to other model types (Wei et al., 2015).

Whilst global methods explain the workings of the overall model, they remain limited in explaining individual predictions. Here methods such as local surrogate models (LIME) (Ribeiro et al., 2016) and SHAP (Lundberg and Lee, 2017) come into play. LIME fits an interpretable surrogate model locally. The local models only explain individual predictions and are not required to be a good global approximation of the model. Whilst LIME has been considered a promising method, it still has

a number of problems, especially for tabular data. For instance, hyperparameter tuning can be challenging with opposite findings being possible depending on the chosen smoothing kernel width. Another issue is the instability of LIME's explanations, which means that repeating the same explanation can lead to significantly different results (Molnar, 2019).

SHAP is a flexible technique which can be used for global interpretation and to explain individual predictions. It has a strong theoretical foundation in game theory and uses the concept of allocating optimal credits based on Shapley values to estimate the importance of features. SHAP force plots provide an intuitive visualisation of how different features affect an individual prediction. One advantage of SHAP for global interpretation, over the Gini and permutation feature importance assessment methods, is that SHAP not only informs about the importance of features but also their relationship with the output. Additionally, SHAP's predictions are fairly distributed among feature values. These factors are critical in guaranteeing trust in the method (Molnar, 2019).

In this work Gini, permutation, and SHAP feature importance assessment were used for global interpretation of the developed models. SHAP was also used to give examples on how individual predictions can be explained locally. The workflow outlined here is summarised by the right-hand side of Fig. 1, which illustrates the workflow of the interpretability analysis and how it interacts with the model development stage described in Sections 2.1 to 2.3.

For interpretation, the results of the analysis were illustrated in graphical form. Global methods were illustrated using bar charts by showing the importance scores across all ten model outputs/targets in the same figure which maximises the information contained per figure. Results were then systematically compared between model types and feature importance assessment methods. As previously mentioned, SHAP also lends itself to the explanation of individual predictions and can globally illustrate the relationship between features and model outputs. Hence additional figures were created to illustrate these aspects.

3. Results and discussion

3.1. Selection of optimal preprocessing steps

As data preparation and preprocessing is a key stage in ML model development, various methods were compared as outlined in Section 2.2. The predictor and target preprocessing methods ALL + DROP and TAR-RF performed best for all but one model type. Predictor methods CONT + DROP and CONT + MEAN which did not include categorical variables performed significantly worse than their counterparts which included categorical variables. This suggests that the use of meaningful categorical variables can noticeably increase the models' prediction accuracy. Mean imputation has frequently been employed due to its simplicity (Hastie et al., 2009). However, comparing mean imputation (CONT + MEAN and ALL + MEAN) to simply dropping the row if missing data was present (CONT + DROP and ALL + DROP) showed that dropping the row generally led to a higher R^2 . These findings indicate that other methods, such as dropping rows with missing data or imputation by a submodel, also need to be considered to ensure optimal model performance.

Target preprocessing method TAR-RF outperformed its alternatives by a considerable margin, which means the added complexity from developing the RF submodels was well justified. By leaving the test data untouched, an unbiased evaluation of the model was possible whilst maximising the models' performance by avoiding missing samples in the training set. The performance of all methods is shown in the supplementary materials.

3.2. Model optimisation and comparison

Initially, the performance between the out of the box models with optimal preprocessing choices and optimised models was compared by

considering the average 5-fold cross-validated R^2 scores shown in Table 1. It was found that RF and XGBoost did not benefit from hyperparameter optimisation, as the out of the box models were found to perform best. The performance of all other model types improved after hyperparameter optimisation as compared to the out of the box models. For instance, the R^2 of GBR slightly increased from 0.87 to 0.90, whereas hyperparameter optimisation had a more substantial effect on the ANN model which showed an increase from 0.63 to 0.81. Generally, optimisation was shown to have a relatively minor effect on ensemble and tree-based methods. However, wherever computational demands during model training are of no concern, hyperparameter optimisation can be an effective mean to obtain the best possible model structure.

In addition to the average 5-fold cross-validated R^2 scores, Table 1 also shows the prediction performance of individual submodels predicting the ten outputs modelled as part of this work. Similarly, the prediction performance in terms of the 5-fold cross-validated RMSE values is shown in brackets. All optimisation results shown in Table 1 used the best preprocessing steps determined during the previous model development stage, as described in Section 3.1. From these results it becomes clear that, overall, GBR was the best performing model type. It achieved the highest performance for seven and eight out of ten outputs in terms of R^2 and RMSE respectively. RF, XGBoost, AdaBoost, ANN, and the SL model were found to perform at an acceptable level with mean R^2 scores > 0.80. Particularly, the SL model had a similar performance to GBR. Whilst ANN models performed exceptionally well in previous works (Ascher et al., 2022b), they were outperformed here. As ANNs require large amounts of data, a potential reason for this is the relatively limited amount of data available to train the model, despite the data set used in this study being still larger than the ones used in previous works (Baruah et al., 2017; Kardani et al., 2021; Zhao et al., 2021).

The SVM model performed well for a few outputs, such as the syngas yield and LHV, but very poorly for other outputs such as the char yield

Table 1

5-fold cross-validated R^2 scores and RMSE (shown in brackets) for all model outputs and model types. The highest scoring model types are shown in bold for each output.

	RF	GBR	XGBoost	AdaBoost	SVM	ANN	Super Learner
N ₂ [vol.% db]	0.96 (4.47)	0.97 (4.08)	0.97 (4.19)	0.96 (4.19)	0.24 (19.7)	0.96 (4.33)	0.97 (4.09)
H ₂ [vol.% db]	0.86 (5.13)	0.89 (2.53)	0.91 (2.89)	0.88 (2.68)	0.58 (5.52)	0.83 (3.36)	0.88 (2.61)
CO [vol.% db]	0.94 (3.90)	0.94 (2.36)	0.82 (2.35)	0.95 (2.14)	0.53 (6.47)	0.93 (2.53)	0.94 (2.10)
CO ₂ [vol.% db]	0.94 (2.82)	0.94 (1.57)	0.79 (1.61)	0.94 (1.57)	0.74 (3.41)	0.88 (2.25)	0.93 (1.59)
CH ₄ [vol.% db]	0.83 (1.58)	0.90 (0.85)	0.77 (1.06)	0.88 (0.93)	0.76 (1.30)	0.84 (1.02)	0.89 (0.87)
C ₂ H _n [vol.% db]	0.77 (0.82)	0.87 (0.40)	0.64 (0.51)	0.78 (0.50)	0.69 (0.61)	0.64 (0.61)	0.85 (0.43)
LHV [MJ/Nm ³]	0.93 (0.79)	0.94 (0.60)	0.94 (0.63)	0.92 (0.69)	0.90 (0.78)	0.88 (0.78)	0.94 (0.60)
Tar [g/Nm ³]	0.59 (15.94)	0.76 (10.29)	0.85 (13.73)	0.41 (14.49)	0.09 (20.47)	0.62 (12.2)	0.75 (9.98)
Gas yield [Nm ³ /kg wb]	0.93 (0.26)	0.96 (0.20)	0.91 (0.25)	0.92 (0.29)	0.90 (0.31)	0.72 (0.45)	0.95 (0.22)
Char yield [g/kg wb]	0.83 (25.11)	0.83 (21.69)	0.76 (21.27)	0.82 (21.92)	0.14 (49.57)	0.76 (26.47)	0.82 (24.84)
Mean	0.86 (6.08)	0.90 (4.46)	0.84 (4.85)	0.85 (4.94)	0.56 (10.81)	0.81 (5.40)	0.89 (4.73)

and syngas' N₂ content. Interestingly, the syngas' N₂ content was an output which was learned extremely well by all other model types with R² scores > 0.96. In contrast, SVM only achieved a R² of 0.24. This suggests that, without extensive further optimisation, SVM is not a suitable method for modelling some of the studied gasification outputs based on the data set provided in this study. When considering the strong out of the box performance of tree-based model such as GBR and RF, it becomes hard to justify the optimisation of SVM models which requires the careful selection of an appropriate kernel function and other hyperparameters.

Fig. 2 shows scatter plots of known target values against model predictions for all optimised GBR submodels. Data points in yellow represent the training set, whereas the test set is shown in blue. The outputs which were modelled well with a high R² and low RMSE, as shown in Table 1, showed little dispersion. Syngas yield and LHV are two examples of outputs with a high prediction accuracy. Other outputs, such as the tar content in the produced syngas, showed a much larger dispersion. When considering test data across all outputs, most predictions have an error of <10 % and very few predictions have an error >20 %, with most outliers being constrained to the syngas tar concentration and char yield submodels. This has been illustrated graphically in the supplementary materials.

A like-for-like comparison of this work to existing studies is not always possible, due to vastly different aims and assumptions. However, in a previous work by the authors an ANN model framework was developed on the same data set (Ascher et al., 2022a). By directly comparing the averaged R² score (averaged across all model outputs) of this previous work, an improvement from 0.86 to 0.90 was achieved by the best performing GBR model of this work. An ANN model suitable for a range of different fluidised bed gasifier bed materials was developed by Serrano et al. (Serrano et al., 2020). Their model achieved R² scores ranging from 0.57 to 0.98 for a range of different outputs (e.g. gas composition in terms of CO₂, CO, CH₄, H₂ and gas yield). When looking at the mean absolute percentage error (MAPE) achieved in their work, there were significant errors ranging from 9.18 % to 38.91 %. The best performing model developed in our work compares favourably with most MAPEs being around or lower than 10 %. Furthermore, our model is also suitable for a wider range of gasifiers, such as fixed-bed gasifiers.

Whilst GBR was found to be the best performing model type in this work, a range of model types were found to be suitable in literature and no single model type appears to be dominant (Ascher et al., 2022b). Elmaz et al. trained polynomial regression, SVM, ANN, and decision tree models on data from an in-house gasifier fed with pine cones and wood pellets (Elmaz et al., 2020). They found decision trees and ANN to be the preferable model types. Despite their data set being more homogenous than the one used in this study, their decision tree models achieved a lower test performance than the ones studied in our work with R² = 0.81–0.94.

Sun et al. took an alternative approach for model optimisation than the one proposed in our work (Sun et al., 2022). By employing particle swarm optimisation, they developed an ANN model for the prediction of syngas yield, gas species concentrations, and char yield with an excellent test performance of R² = 0.97. However, the authors highlighted that the model had only been trained on data from pine wood gasification and increasing the size of the data set by incorporating a wider range of feedstocks and gasification conditions was deemed important to improve the model's applicability.

The model developed in this work has the potential to simplify general gasification process design. By using a varied data set for model training a large range of different gasification systems and feedstocks could be optimised and compared. Furthermore, the model's predictions could be used in a more holistic system modelling context. For instance, predictions could be directly used for LCSA or techno-economic analysis (TEA).

3.3. Interpretability analysis

Initially, the global interpretability of models was studied. The GBR model was taken as the baseline model as it was found to be the best performing model. It was then compared to the three next best performing model types, namely AdaBoost, RF, and XGBoost. It must be noted that at the time of writing the SHAP method was not supported for AdaBoost, hence only Gini and permutation importance were computed for this model type.

Fig. 3(a) and (b) shows the Gini feature importance of the GBR and RF models across all ten outputs. The x-axis shows the importance scores, whereas the y-axis shows the predictors used to train the models. Submodels predicting the different outputs can be differentiated by their colours. RF is shown for comparison purposes, because as a bagging model it operates differently from the GBR model. However, other model types are also discussed. Additional figures for other model types and feature importance assessment methods are shown in the supplementary materials folder of the GitHub repository shared in the Data Availability section. The feedstock's particle size and choice of gasifying agent were found to be the key predictors across all four studied model types, with the particle size being the top predictor for GBR, RF, and AdaBoost. The top ten predictors account for most of the variation in the GBR and RF models' predictions. This is illustrated by a drop of over 50 % in the combined Gini importance scores from the 10th to 11th most important predictor.

Looking at the ten most influential predictors, it can be seen that they are identical for GBR and RF, with predictors such as the temperature, ER, and proximate and ultimate composition strongly affecting the models' predictions. Some minor differences for AdaBoost were apparent, whereas XGBoost showed some more major deviations. Where most submodels generally substantially contributed to a feature's importance for GBR, RF, and AdaBoost, this is not the case for XGBoost. For this model type, a feature's combined importance was often heavily dominated by a few submodels. For instance, the score of XGBoost's top predictor gasifying agent oxygen was largely made up by the three submodels predicting the syngas yield, syngas CO₂ content, and syngas LHV.

Comparing Gini to permutation importance for RF showed few differences, with the top three predictors remaining the same. Similarly, the top predictors for GBR remained nearly unchanged when comparing Gini to permutation importance, with only catalyst usage displacing oxygen as a gasifying agent in the top ten as shown by Fig. 3(c). The use of a catalyst was found to impact the prediction of the syngas' tar content most strongly. This is in good agreement with literature, as many studies used a catalyst with the goal of reducing the tar content in the produced syngas (De Andrés et al., 2011; Luo et al., 2012).

Whilst the char yield submodel only contributed marginally to the Gini importance findings, it dominated other submodels based on permutation importance assessment. For permutation importance, the gasification temperature was found to be the most important predictor for GBR, with more than two thirds of the combined score coming from the char yield prediction submodel.

In general, when looking at the combined feature importance of all submodels, all studied model types were in good agreement for permutation-based feature importance. However, the overall importance of individual predictors could result from very different submodels. For instance, the particle size, which was found to be the most important predictor for three out of four model types, had significant contributions to its importance from most outputs for GBR and XGBoost. In contrast, for AdaBoost its importance was heavily dominated by the outputs CO and C₂H₄. When looking at RF, N₂, C₂H₄, and char yield were the outputs which mostly contributed to the particle size's importance as a predictor.

The feature importance results yielded by the SHAP method are shown by Fig. 3(d) for GBR. Top predictors were found to be similar to Gini and permutation feature importance. Where permutation

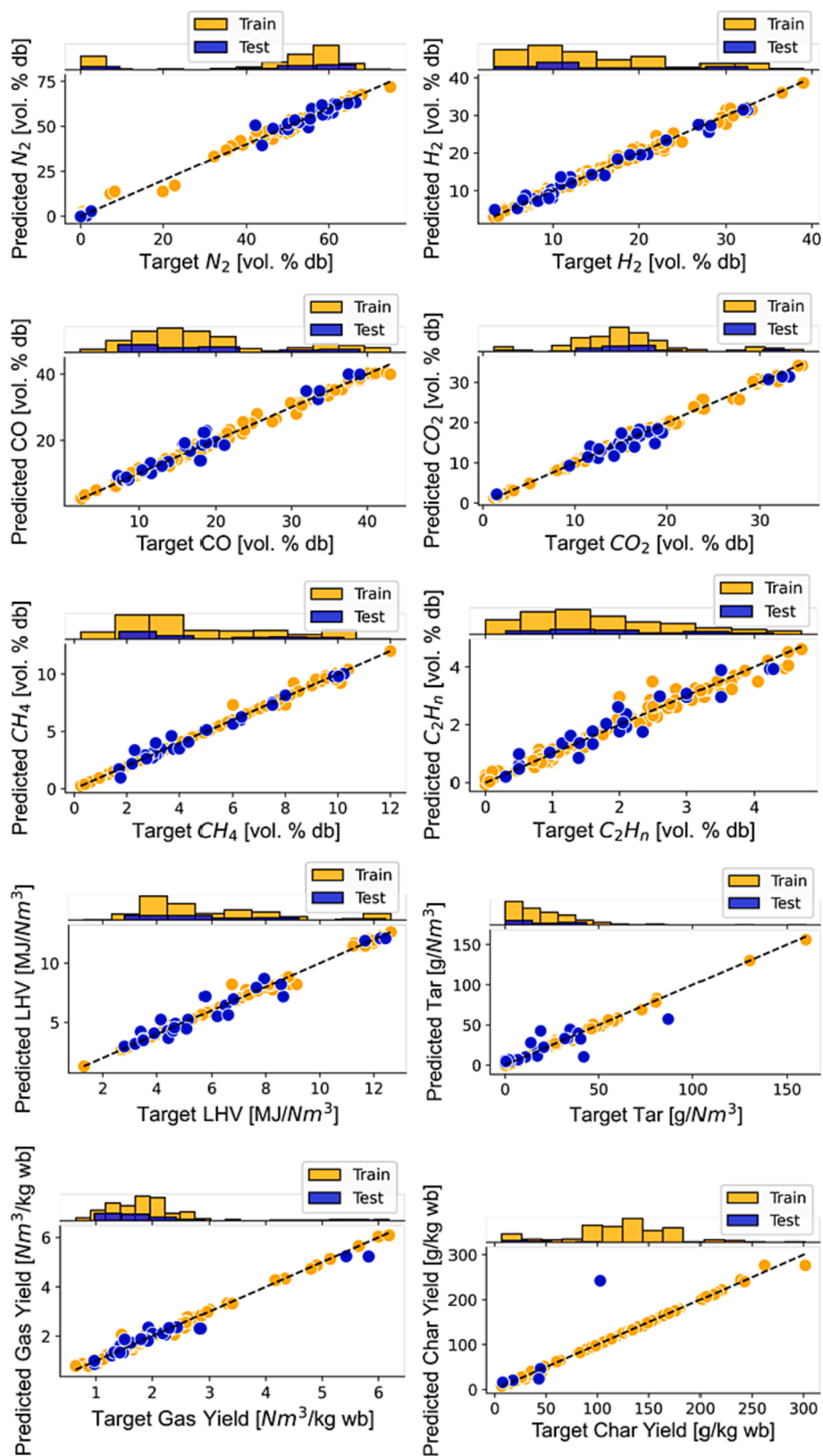


Fig. 2. Scatter plots of target values vs predictions for all 10 optimised GBR submodels. The figures show both the training (yellow) and test (blue) performance of each model. The black dashed line indicates perfect predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

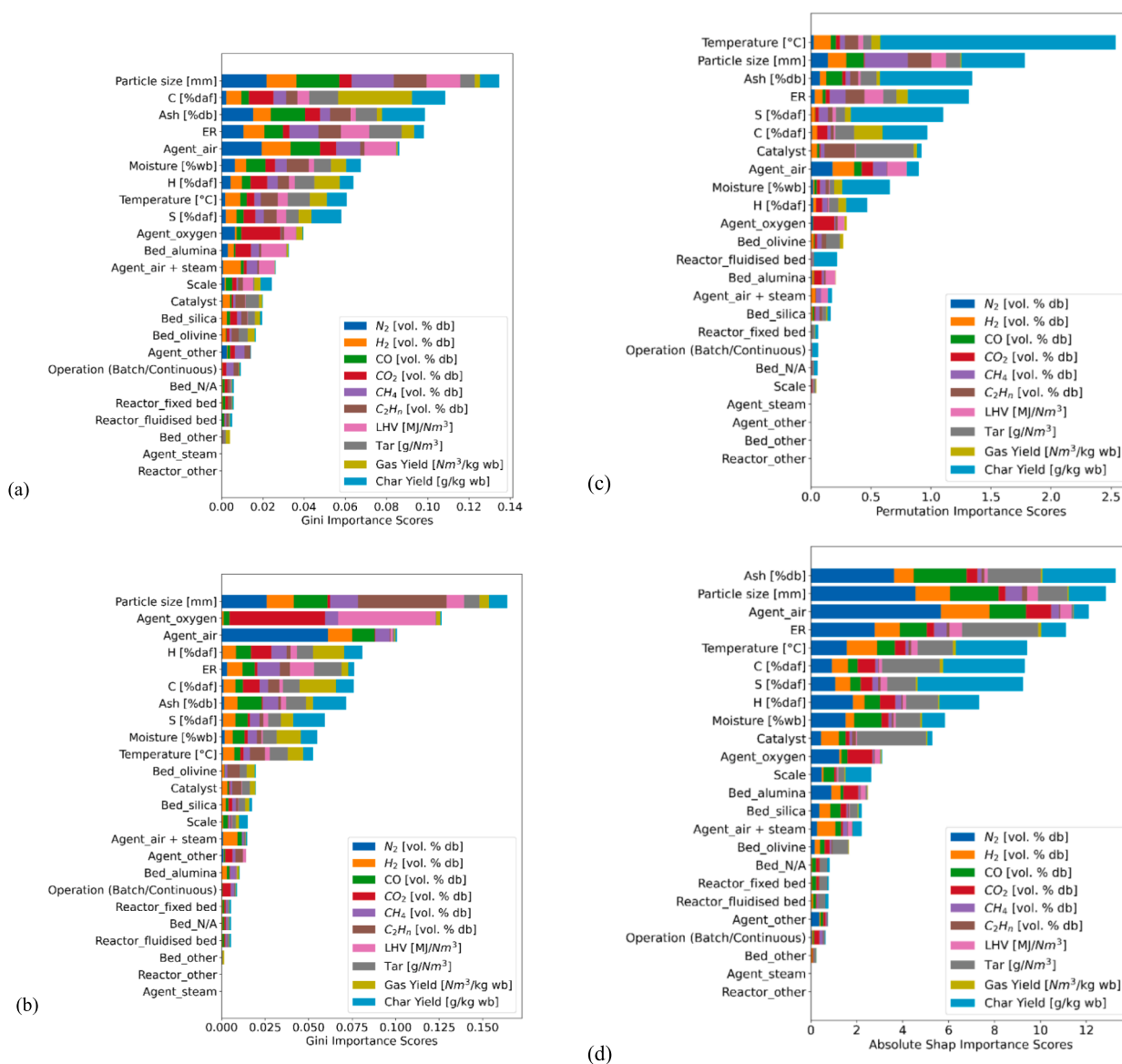


Fig. 3. Gini feature importance of (a) gradient boosting models and (b) random forest models, as well as (c) permutation and (d) SHAP feature importance analysis of gradient boosting models. Importance scores are shown on the x-axis. The y-axis shows the predictors used to train models for the 10 considered outputs. A predictor's total score is made up of the importances from the different submodels (i.e. model outputs) which are illustrated by the different colours shown in the legend.

importance was heavily dominated by individual outputs, SHAP-based feature importance scores showed more balanced contributions. A finding worth highlighting is that whilst the N₂ content submodel contributed to a similar level as other submodels to the importance of air as a gasifying agent for Gini and permutation feature importance, this contribution was much larger for SHAP feature importance. RF and XGBoost already showed a much larger contribution from this submodel to the importance of air as a gasifying agent for Gini and permutation feature importance. For SHAP feature importance, this dominance was more obvious and more than half of the importance score of air as a gasifying agent resulted from the N₂ content submodel. These findings are in good agreement with literature, where the use of air as an agent has been linked to a diluted syngas with a high N₂ concentration (Lui et al., 2020; Sikarwar et al., 2017).

Even though all three methods generally agreed well with each other, there was some variation in the importance of predictor variables. The most striking difference was that the combined permutation-based

feature importance scores could be heavily dominated by a single output, whilst this was less obvious for the Gini and SHAP-based scores. Gini-based feature importance is generally prone to favour high cardinality features, such as numerical inputs (Breiman, 2001). However, in this work, it was found that some categorical variables (e.g. type of gasifying agent used) were also important.

Whilst Fig. 3 shows the absolute importance of features across all outputs, the nature of their relationships with the outputs remains unexplained. The SHAP method can be used to not only study the importance of features but also their relationships with individual output parameters. Fig. 4(a) shows the absolute importance of features in a similar fashion to Fig. 3, but this time for a single output. In this instance, the syngas yield is shown. Fig. 4(b) not only illustrates the feature importance but also feature effects. Red represents a high feature value, whereas blue represents a low feature value. The further away a point is from the baseline SHAP value of zero, the stronger it effects the output. This way a features relationship with the SHAP value (and in turn the

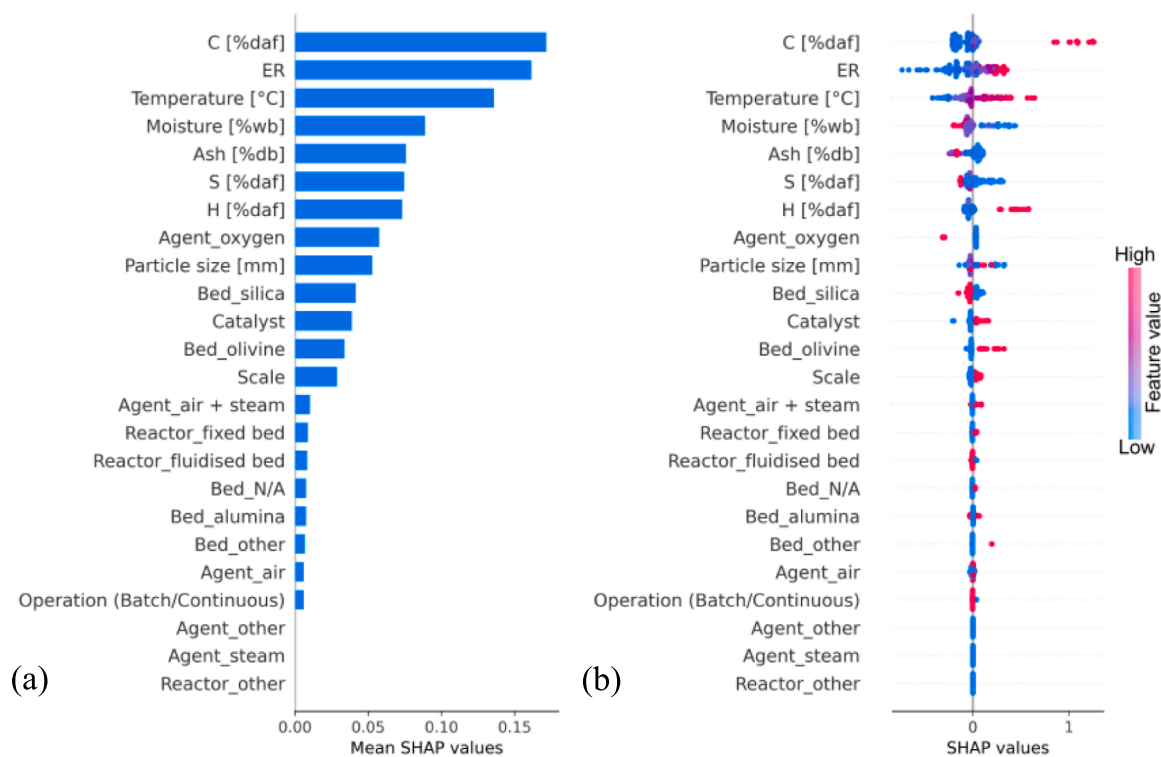


Fig. 4. Feature importance analysis by SHAP method for GBR model predicting the syngas yield. Absolute importance scores are shown by (a), whereas (b) shows the influence of individual predictions on the overall importance scores. Red represents a high feature value (in this case syngas yield), whereas blue represents a low feature value. The further away a point is from the baseline SHAP value of zero, the stronger it effects the output. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predicted output) can be better understood.

The feedstock's carbon content most strongly affected the GBR model's prediction for the syngas yield. From Fig. 5(b) it becomes clear that a low feedstock carbon content led to a minor decrease in syngas yield. A high carbon content led to a large increase in syngas yield. Most data points show a low feature value and are bulked in the negative or near zero SHAP value region. The data points with a high feature value (SHAP value close to 1) correspond to plastic waste and other feedstocks with a very high carbon content. The temperature and ER show a smoother distribution of data points. Again, low temperatures and ERs decreased the syngas yield, whereas high feature values increased the syngas yield. Feature values near the middle of the range (shown in purple) were shown to have little effect on the model's predictions. For the binary variable catalyst usage, it was shown that the use of a catalyst generally had a positive effect on the syngas yield. Figures for the remaining nine model outputs can be found in the [supplementary materials](#) folder of the shared GitHub repository.

Fig. 5 illustrates how the SHAP method can be used to explain individual model predictions. Two example model outputs and feedstocks are shown in the figure. The figures represent an intuitive way to guide the decisions of stakeholders and investors and improve their understanding of how the developed model makes a particular prediction. Fig. 5(a) and (b) show the syngas yield from barley straw gasification for a base case at 800 °C (a) and a high temperature case at 1,000 °C (b). The gasification process underwent continuous operation without a catalyst in a pilot-scale fluidised bed gasifier using silica as a bed material and air as an agent. A temperature of 800 °C, particle size of 4 mm, and an ER of 0.3 were selected for the base case (Fig. 5(a)), as these values are close to the mean of all samples for these variables. The force plots show how individual feature values perturb the model's prediction from a baseline value. For the base case, as shown by Fig. 5(a), most feature values

lowered the predicted syngas yield (shown in blue). Most notably, the barley straw's low carbon and relatively high moisture contents led to a decrease in the model's predicted value. When raising the temperature to 1000 °C as shown by Fig. 5(b), a significant increase in the predicted syngas yield can be noticed (shown in red). The model's predicted syngas yield rose from 1.63 to 2.21 Nm³/kg wb. The high temperature value solely balanced the effects of all feature values lowering the predicted yield.

For Fig. 5(c), the gasifier setup and conditions remained unchanged from the ones previously described (i.e. shown in Fig. 5(a) and (b)), however the feedstock has been changed to MSW. Two variables have been altered for Fig. 5(d), namely steam was used as a gasifying agent instead of air and the feedstock's particle size was reduced from 4 mm to 1 mm. The syngas LHV has been selected as the output of interest in this example. MSW's high feedstock carbon content led to an increase in the model's prediction. In contrast, a high feedstock ash content of 16.82 % db and a particle size of 4 mm led to a reduction in the predicted LHV. Fig. 5(d) reveals that a smaller particle size can increase the predicted LHV. Similarly, using steam as a gasifying agent was found to significantly increase the model's predicted LHV. Interestingly, this was illustrated by the one-hot encoded variable representing the gasifying agent air equalling zero. This could be interpreted as the model understanding that air as a gasifying agent reduced the LHV. The opposite applied to the variable gasifying agent oxygen equalling zero – here the model learned that the gasifying agent not being oxygen adversely affected the syngas' LHV.

In summary, Fig. 5 shows how the SHAP method could be used to communicate a ML model's prediction process to a non-expert audience. This way the ML practitioner can offer a simple explanation of how certain system set-ups effect the outputs which stakeholder or policy-makers may be concerned with. Different system or feedstock choices

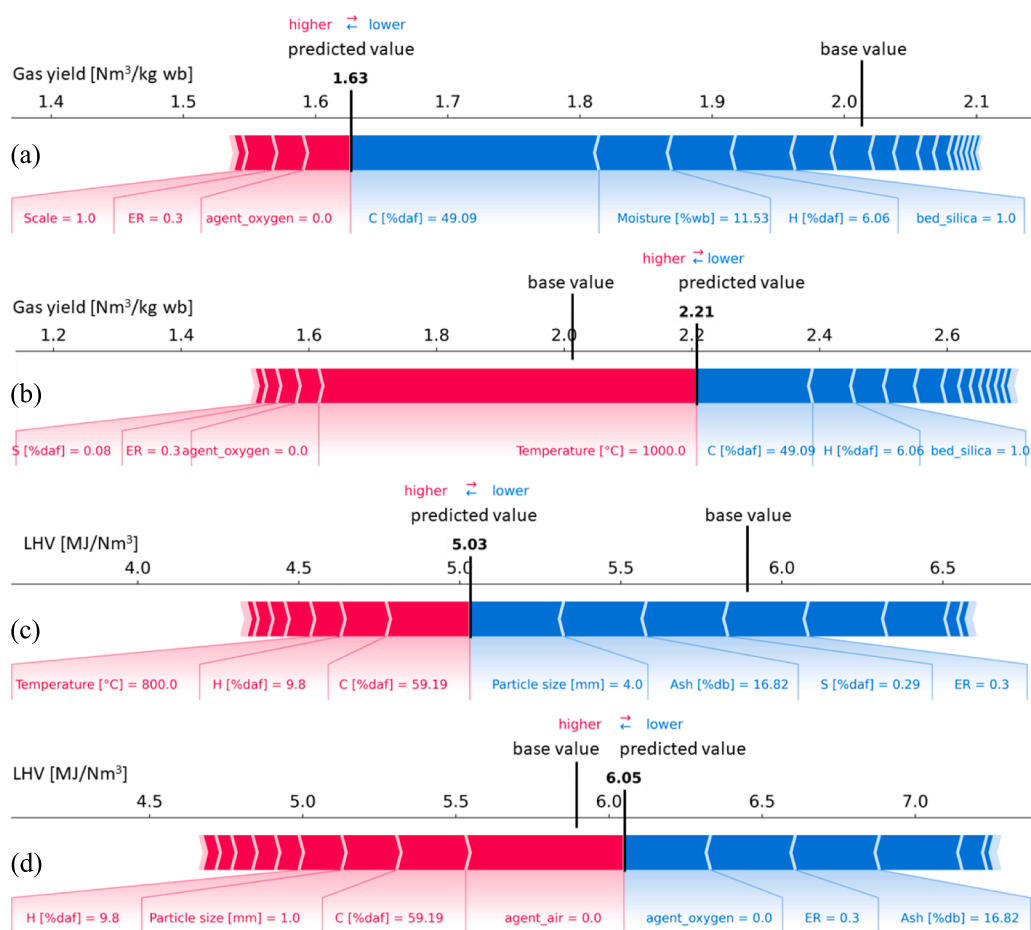


Fig. 5. SHAP explanations of some selected individual predictions. The syngas yield from barley straw gasification is shown for a base case at 800 °C (a) and a high temperature case at 1,000 °C (b). The resulting syngas LHV from MSW gasification is shown for a base case using air as a gasifying agent (c) and an alternative case using steam as a gasifying agent (d). The force plots start at the base value (the average of all predictions). Each predictor (and its corresponding Shapley value) is represented by an arrow which either increases (shown in red) or decreases (shown in blue) the model's predicted value with respect to the base value. A predictor's importance is shown by the size of its arrow, where a larger arrow represents a more important predictor. Ultimately, the model's predicted value is illustrated by the point where the red and blue arrows meet. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

may be studied by creating similar figures for an array of operating conditions and feedstock combinations.

Studies assessing the interpretability of bioenergy ML models are still scarce. Researchers have often looked towards PDPs and Pearson's or Spearman's correlation coefficients (Li et al., 2021a; Yuan et al., 2021; Zhao et al., 2021; Zhu et al., 2019). PDP are an effective and straightforward mean to illustrate the relationship between a predictor variable and model output. However, as each input and output combination requires an individual plot this method can become infeasible, especially for multiple-input multiple-output (MIMO) models like the one presented in this work. For this reason, identifying the importance of features through global methods, to then study the relationship in more detail using PDP, represents a good alternative. Furthermore, confounding factors cannot be captured by PDP.

Another popular interpretability method has been permutation feature importance (Li et al., 2021c; Yuan et al., 2021; Zhao et al., 2021). A key benefit of permutation importance is its easy implementation and the fact that it can be implemented for any model type. However, by randomly shuffling a feature, unrealistic data instances can be created (e.g. feedstock ultimate composition no longer summing to 100 %) which may limit interpretability.

Zhao et al. studied the supercritical water gasification of biomass for hydrogen production (Zhao et al., 2021). By using feature permutation on a RF model, the authors identified the biomass concentration and

temperature to be the two most influential factors effecting hydrogen production. Our work found gasifier temperature to be of medium importance with features such as the gasifying agent and particle size being more important in predicting hydrogen production. However, these factors were not considered in Zhao et al.'s work.

Li et al. developed a GBR model to predict the syngas yield from the hydrothermal gasification of wet waste (Li et al., 2021b). By assessing the feature importance, the authors showed that the model heavily relied on the gasifier temperature to make its predictions. Other factors such as the feedstock composition had a lesser effect. In comparison, in our work the gasifier temperature was found to be important but not multiple times more important than other factors as in Li et al.'s study. However, a direct comparison is difficult as the studied gasification process is significantly different from the ones studied in our work.

A challenge for all types of feature importance assessment are correlated features. As they contain largely the same information, their importance can be split across two or more features, making the correlated features appear less important than they are. This highlights the importance of removing strongly correlated features to ensure that important features do not get lost. Hence, a rigorous predictor selection process is required when feature importance assessment is one of the aims of model development. In this work, this issue has been addressed by computing Pearson's and Spearman's correlation coefficients to subsequently remove strongly correlated features. Finally, the difference

between causation and correlation must be highlighted. Feature importance assessment can only bring to light which features the model uses to make its predictions. Although the behaviour of the model does not necessarily correspond to a real-world process, it may provide one with a good starting point on which variables may significantly affect some of the outputs one is interested in. Additionally, it allows sensitivity checks to confirm whether inputs deemed important by theory are also deemed important by the model.

By combining both global and local methods, a more complete picture of a model's workings can be established. Global models powerfully describe the average behaviour of a ML Model and highlight trends in the data. Local explanations on the other hand are useful for understanding which factors might influence a given gasification system. This provides researchers with an idea on which parameters could be changed to optimise a given output. For instance, the developed method could be used to screen and test for promising process conditions before moving on to more expensive and time-consuming physical experiments. Furthermore, SHAP makes it easy to communicate the entire thought process to stakeholders. However, care needs to be taken when interpreting SHAP results as interpretability methods can be fooled to hide biases (Slack et al., 2020). As such, it is the researcher's responsibility to avoid creating misleading explanations.

4. Conclusions

Gradient boosting regression (GBR) has been found to outperform other model types with a coefficient of determination (R^2) of 0.90 when averaged across all ten model outputs.

Global and local interpretability methods were combined to extract new information from the developed models. This information can be used to guide the decisions of investors and policy makers by increasing their confidence in the models' results. This reduction in uncertainty can in turn promote the uptake of a new technology. The feedstock's particle size and the gasifying agent were amongst the top predictor variables influencing the models.

CRedit authorship contribution statement

Simon Ascher: Conceptualization, Methodology, Software, Writing – original draft. **Xiaonan Wang:** Writing – review & editing. **Ian Watson:** Supervision. **William Sloan:** Funding acquisition. **Siming You:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

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

All data supporting this study are provided in full in the 'Methodology' and 'Results and Discussion' sections, the e-supplementary material of this paper which can be found in the e-version of this paper online, or at <https://github.com/SimonAscher/Interpretable-Machine-Learning-to-Model-Biomass-and-Waste-Gasification>, an open-source online repository.

Acknowledgements

The authors would like to acknowledge the financial support from the Engineering and Physical Sciences Research Council (EPSRC) Studentship and Programme Grant (EP/V030515/1). Siming You would like to thank the financial support from the Royal Society International Exchange Scheme (EC\NSFC\211175).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.biortech.2022.128062>.

References

- Ascher, S., Sloan, W., Watson, I., You, S., 2022a. A comprehensive artificial neural network model for gasification process prediction. *Appl. Energy* 320, 119289. <https://doi.org/10.1016/j.apenergy.2022.119289>.
- Ascher, S., Watson, I., You, S., 2022b. Machine learning methods for modelling the gasification and pyrolysis of biomass and waste. *Renew. Sustain. Energy Rev.* 155, 111902. <https://doi.org/10.1016/j.rser.2021.111902>.
- Baruah, D., Baruah, D.C., Hazarika, M.K., 2017. Artificial neural network based modeling of biomass gasification in fixed bed downdraft gasifiers. *Biomass and Bioenergy* 98, 264–271. <https://doi.org/10.1016/j.biombioe.2017.01.029>.
- BEIS, 2021. Biomass Policy Statement.
- Breiman, L., 2001. *Random Forests*. *Mach. Learn.* 45, 5–32.
- S.L. Brunton, J.N. Kutz Data-driven science and engineering: machine learning, dynamical systems, and control 2017 10.1109/mcs.2021.3076544.
- De Andrés, J.M., Narros, A., Rodríguez, M.E., 2011. Air-steam gasification of sewage sludge in a bubbling bed reactor: Effect of alumina as a primary catalyst. *Fuel Process. Technol.* 92, 433–440. <https://doi.org/10.1016/j.fuproc.2010.10.006>.
- Elmaz, F., Yücel, Ö., Mutlu, A.Y., 2020. Predictive modeling of biomass gasification with machine learning-based regression methods. *Energy* 191, 116541. <https://doi.org/10.1016/j.energy.2019.116541>.
- Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Modell.* 160, 249–264. [https://doi.org/10.1016/S0304-3800\(02\)00257-0](https://doi.org/10.1016/S0304-3800(02)00257-0).
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning*, Second. ed, Springer Series in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-84858-7>.
- Kardani, N., Zhou, A., Nazem, M., Lin, X., 2021. Modelling of municipal solid waste gasification using an optimised ensemble soft computing model. *Fuel* 289, 119903. <https://doi.org/10.1016/j.fuel.2020.119903>.
- Kégl, B., 2013. The return of ADABOOST.MH: multi-class Hamming trees. arXiv <https://doi.org/10.48550/ARXIV.1312.6086>.
- L. Lacrosse M. Englisch K. Danner H. Stokes The role of bioenergy in the clean energy transition and sustainable development – lessons from developing countries 2021.
- Li, J., Pan, L., Suvarna, M., Tong, Y.W., Wang, X., 2020. Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning. *Appl. Energy* 269, 115116. <https://doi.org/10.1016/j.apenergy.2020.115116>.
- Li, J., Pan, L., Suvarna, M., Wang, X., 2021a. Machine learning aided supercritical water gasification for H₂-rich syngas production with process optimization and catalyst screening. *Chem. Eng. J.* 426, 131285. <https://doi.org/10.1016/j.cej.2021.131285>.
- Li, J., Suvarna, M., Pan, L., Zhao, Y., Wang, X., 2021b. A hybrid data-driven and mechanistic modelling approach for hydrothermal gasification. *Appl. Energy* 304, 117674. <https://doi.org/10.1016/j.apenergy.2021.117674>.
- Li, J., Zhu, X., Li, Y., Tong, Y.W., Ok, Y.S., Wang, X., 2021c. Multi-task prediction and optimization of hydrochar properties from high-moisture municipal solid waste: Application of machine learning on waste-to-resource. *J. Clean. Prod.* 278, 123928. <https://doi.org/10.1016/j.jclepro.2020.123928>.
- Lui, J., Chen, W.-H., Tsang, D.C.W., You, S., 2020. A critical review on the principles, applications, and challenges of waste-to-hydrogen technologies. *Renew. Sustain. Energy Rev.* 134, 110365.
- Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: *Advances in Neural Information Processing Systems (NIPS 2017)*.
- Luo, S., Zhou, Y., Yi, C., 2012. Syngas production by catalytic steam gasification of municipal solid waste in fixed-bed reactor. *Energy* 44, 391–395. <https://doi.org/10.1016/j.energy.2012.06.016>.
- Molnar, C., 2019. *Interpretable Machine Learning – A Guide for Making Black Box Models Explainable*.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *Knowl. Discov. Data Min.* doi 10 (1145/2939672), 2939778.
- Serrano, D., Castelló, D., 2020. Tar prediction in bubbling fluidized bed gasification through artificial neural networks. *Chem. Eng. J.* 402, 126229. <https://doi.org/10.1016/j.cej.2020.126229>.
- Serrano, D., Golphour, I., Sánchez-Delgado, S., 2020. Predicting the effect of bed materials in bubbling fluidized bed gasification using artificial neural networks (ANNs) modeling approach. *Fuel* 266, 117021. <https://doi.org/10.1016/j.fuel.2020.117021>.
- Sikarwar, V.S., Zhao, M., Fennell, P.S., Shah, N., Anthony, E.J., 2017. Progress in biofuel production from gasification. *Prog. Energy Combust. Sci.* 61, 189–248. <https://doi.org/10.1016/j.pecc.2017.04.001>.
- Slack, D., Hilgard, S., Jia, E., Singh, S., Lakkaraju, H., 2020. Fooling LIME and SHAP, in: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. pp. 180–186. <https://doi.org/10.1145/3375627.3375830>.
- C. Sun L. Ai T. Liu The PSO-ANN modeling study of highly valuable material and energy production by gasification of solid waste: an artificial intelligence algorithm approach 2022 *Biorefinery Biomass Convers* 10.1007/s13399-022-02342-2.
- Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wu, Z., Chen, H., Naqvi, S.R., 2020. Prediction of Bio-oil Yield and Hydrogen Contents Based on Machine Learning

- Method: Effect of Biomass Compositions and Pyrolysis Conditions. *Energy & Fuels* 34, 11050–11060. <https://doi.org/10.1021/acs.energyfuels.0c01893>.
- Van Der Laan, M.J., Polley, E.C., Hubbard, A.E., 2007. Super learner. *Stat. Appl. Genet. Mol. Biol.* 6 <https://doi.org/10.2202/1544-6115.1309>.
- Wei, P., Lu, Z., Song, J., 2015. Variable importance analysis: A comprehensive review. *Reliab. Eng. Syst. Saf.* 142, 399–432. <https://doi.org/10.1016/j.res.2015.05.018>.
- Yuan, X., Suvarna, M., Low, S., Dissanayake, P.D., Lee, K.B., Li, J., Wang, X., Ok, Y.S., 2021. Applied Machine Learning for Prediction of CO₂ Adsorption on Biomass Waste-Derived Porous Carbons. *Environ. Sci. Technol.* 55, 11925–11936. <https://doi.org/10.1021/ACS.EST.1C01849>.
- Zhang, T., Cao, D., Feng, X., Zhu, J., Lu, X., Mu, L., Qian, H., 2022. Machine learning prediction of bio-oil characteristics quantitatively relating to biomass compositions and pyrolysis conditions. *Fuel* 312, 122812. <https://doi.org/10.1016/j.fuel.2021.122812>.
- Zhao, S., Li, J., Chen, C., Yan, B., Tao, J., Chen, G., 2021. Interpretable machine learning for predicting and evaluating hydrogen production via supercritical water gasification of biomass. *J. Clean. Prod.* 316, 128244 <https://doi.org/10.1016/j.jclepro.2021.128244>.
- Zhu, X., Li, Y., Wang, X., 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. *Bioresour. Technol.* 288, 121527 <https://doi.org/10.1016/j.biortech.2019.121527>.