RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2023, **13**, 802

Predicting the high heating value and nitrogen content of torrefied biomass using a support vector machine optimized by a sparrow search algorithm†

Liu Xiaorui, [a] Yang Jiamin[a] and Yuan Longji*[b]

A support vector machine (SVM) model with RBF kernel function combined with sparrow search algorithm (SSA) optimization was developed to predict the HHV and nitrogen content (No) values of torrefied biomass based on the feedstock properties and torrefaction conditions. Results showed that SSA optimization significantly improved the prediction performance of the SVM model for both HHV and No. A coefficient of determination (R^2) larger than 0.91 was achieved when the SSA-SVM model was implemented, and the values of RMSE were also fairly acceptable. The agreement between experimental data and SSA-SVM predicted values demonstrated the high predictive precision of the model. This study provides a reference for the utilization of torrefied biomass in solid fuels and the design of torrefaction facilities.

Received 31st October 2022
Accepted 14th December 2022

DOI: 10.1039/d2ra06869a

rsc.li/rsc-advances

Introduction

Global warming caused by the excessive consumption of fossil fuels, which produce large amounts of greenhouse gas (mainly CO_2) emissions, has led to the diminishment of snow and ice, as well as a global average sea level rise.[1] Due to this, the utilization of sustainable and renewable energy sources such as solar, wind, and biomass energy, *etc.* becomes increasingly important. Among these renewable sources, biomass, which is the only carbon-based material, has attracted increasing attention. Different from solar and wind energy, biomass can not only be used for power generation and heat supply but can also be converted into gaseous, liquid and solid products with zero CO_2 emissions through its life cycle.[2] However, some inherent disadvantages of biomass prevent its large-scale utilization, such as the high moisture content, low calorific value and low energy density which would result in the high costs for biomass collection, storage and transportation.[3]

Torrefaction, which is traditionally performed at 200–300 °C in inert atmosphere is a promising pretreatment technique to overcome the above mentioned shortcomings of raw biomass.[4–6] Few studies also include an oxidative atmosphere such as air, steam, flue gas, *etc.* into the definition of torrefaction.[7,8] In order to estimate the torrefaction process, characterization of torrefied biomass properties is necessary.[9] Higher heating value (HHV), an important parameter for designing the biomass conversion facilities, is either experimentally tested by a calorimeter bomb or mathematically calculated using Channiwala and Parikh's correlation based on the proximate and ultimate analysis results.[4,10] These characterizations always require repetitive experiments and subsequently instrumental analysis or mathematical method, consuming lots of time, costs and manpower. Therefore, developing a reliable method to predict the properties of torrefied biomass based on those of feedstock without various tests and experiments is of great value to save time, costs and manpower.

Machine Learning (ML), a subset of artificial intelligence (AI), has a strong ability to deal with the multi-dimensional and non-linear problems involving classification and regression with the superiorities of time-saving and high prediction accuracy.[11] Due to these advantages, ML can be applied to the prediction of torrefied biomass properties to avoid repetitive experiments. For torrefied biomass, the artificial neural network (ANN) models were most frequently used to predict the yields,[12,13] the CHO contents and HHV,[9] and the exergy[14] of torrefied biomass. In addition to the ANNs, several other models have also been developed to predict the properties of torrefied product. García *et al.*[15,16] predicted the HHV of torrefied biomass using support vector machine models (SVMs) combined with particle swarm optimization (PSO) or simulated annealing (SA) optimization. Onsree *et al.*[17,18] compared the accuracy of kernel ridge regression (KRR), gradient tree boosting (GTB), ANN, SVM and random forest (RF) models in predicting the yields. Leng *et al.*[19] employed extreme gradient boosting (XGB), RF, SVM, and multilayer perceptron (MLP) algorithms to predict the distribution of three-phase product. These studies indicated that ML algorithms are capable of predicting the properties of torrefied biomass especially the yield and HHV. However, it is worth noting that no single algorithm was optimal for all problems since every system had its unique data structure.[20]

[a]*School of Mine, China University of Mining and Technology, 221116 Xuzhou, China. E-mail: liuxiaorui214@cumt.edu.cn*

[b]*School of Low-Carbon Energy and Power Engineering, China University of Mining and Technology, 221116 Xuzhou, China. E-mail: yuanlongji@cumt.edu.cn*

† Electronic supplementary information (ESI) available: Table S1: dataset; Table S2: dataset description; Table S3. Pearson correlation coefficient between any two features. See DOI: https://doi.org/10.1039/d2ra06869a

Moreover, application of ML algorithms in biomass torrefaction is fairly rare and existing literatures mainly focused on the prediction of yield and HHV, whereas other properties of torrefied biomass were seldomly concerned.

It is well known that NOx is the main gaseous pollutants during biomass utilization due to the higher conversion rate of fuel-N than occurs with coal.[21] Besides, when biomass is pyrolyzed and gasified, a considerable proportion of fuel-N was released in the forms of NH_3 and HCN which are harmful to the environment and human health.[22] These nitrogen containing pollutants were mainly originated from the conversion of the nitrogen element in the feedstocks. Therefore, the nitrogen content of torrefied biomass should also be deservedly attended. However, to the authors' knowledge, the prediction of nitrogen content of torrefied biomass has not been reported.

As reviewed above, there are still large gaps in the prediction of torrefied biomass properties using machine learning method. In this study, two important fuel parameters involving the HHV and nitrogen content of torrefied biomass were estimated by the support vector machine (SVM) with RBF kernel function. To improve its prediction accuracy, the SVM model was further optimized by sparrow search algorithm (SSA) technique to develop a novel SSA-SVM model. To date, this is the first study to implement the SSA-SVM model to predict the properties of torrefied biomass.

Methods

Dataset collection and pre-processing

497 data points were extracted from 66 peer-reviewed publications with respect to the traditional torrefaction which was performed at 200–300 °C in inert atmosphere (N_2, He, Ar and anoxic environment) to create a dataset. The detailed information of the dataset was given in the ESI (Table S1†). Most of the data points were directly extracted from the text and the tables of the papers and the corresponding ESI.† While for those data which were not directly listed, WebPlotDigitizer tool (**https://apps.automeris.io/wpd/**) was employed to extract the necessary data from the figures. The proximate analysis results of the raw biomass involving volatile matter (VM, wt%), fixed carbon (FC, wt%) and ash (Ash, wt%) contents, the ultimate analysis results (C, H, O, Ni, wt%) and torrefaction conditions involving temperature (Temp, °C) and duration time (time, min) were collected as input features. The HHV (MJ kg^{-1}) values and nitrogen content (No) of the torrefied biomass were assigned to the targets.

Then, correlation analysis was performed and the linear dependency among all the input and output variables was measured by Pearson's correlation coefficient (PCC).[23]

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\sum_{i=1}^{n}(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

where r is the value of PCC ranging from -1 to 1, where 0 means no linear correlation, and a negative or positive value indicates a negative or positive correlation, respectively. The greater the absolute value of r, the stronger the linear correlation; $\bar{x}$ and $\bar{y}$ are the means of the input feature and output target, respectively. The two features were considered to be strongly correlated to each other when the absolute value of r is greater than 0.7, and one of them will be excluded.[24]

To obtain a uniform range among the variables, the features and targets were normalized with Z-score standardization using eqn (2):

$$x_i^* = \frac{x_i - \mu}{s} \quad (2)$$

where x_i is the value of input feature i; x_i^* is the normalized value of initial x_i; μ is the mean value of x_i, and s represents its standard deviation.

Then the preprocessed dataset was randomly divided into training and testing subsets at a ratio of 80% to 20% for the evaluation of the developed models.

Training models

SVM model. SVM is a supervised machine learning model that originally developed for two-group classification.[15] However, it was soon extended to work for continuous outcomes with small samples, high-dimensions, and nonlinearity which is called support vector regression (SVR). When dealing with complex and nonlinear regression problems, the kernel function is particularly useful because it can map the features into high dimensional space to linearly separable variables[25] and greatly facilitate the computations.[26]

SVR can be formulated as an optimization problem (eqn (3)) to minimize the norm of the weight vector (ω) with some slack variables (ξ_i, ξ_i^*) introduced to increase the tolerance of regression error.[27]

$$\min_{\omega,b,\xi_i,\xi_i^*} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*), \xi_i, \xi_i^* \geq 0, i = 1, \ldots, m \quad (3)$$

The above optimization (eqn (3)) can be done by solving its dual problem (eqn (4)) with a Lagrange dual formulation.

$$\max \sum_{i=1}^{m} y_i(\alpha_i^* - \alpha_i) - \varepsilon\sum_{i=1}^{m} y_i(\alpha_i^* + \alpha_i) - \frac{1}{2}\sum_{i,j=1}^{m} y_i(\alpha_i^* + \alpha_i) \times (\alpha_i^* - \alpha_i){x_i}^T x_j \quad (4)$$

where C is the penalty term that determines the trade-off between the misclassifications of the training data and the margin width. α_i and α_i^* are the Lagrange multipliers, and ε is the tolerance of margin.

In this study, SVR model with the RBF kernel function (eqn (5)) was employed to predict HHV and No.

$$K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2), \gamma > 0 \quad (5)$$

where γ defines the effect of a single sample on the entire classification hyperplane. Then, the SVR model with the RBF kernel function can be described as:

$$f(x) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (6)$$

where b refers to the bias term.

SSA-SVM model. Penalty factor C and kernel function are the two important hyper-parameters that affect the SVM model.[28] The selection of C and kernel function parameter can considerably influence the regressor results. Traditional SVMs are not sensitive to outliers and are easy to fall into local optimal solutions. Thus, sparrow search algorithm (SSA), which is a novel swarm intelligence optimization algorithm proposed by Xue *et al.*[29] in 2020 based on the behavior of sparrows foraging and evading predators was employed to further optimize the SVM model. Meanwhile, the SSA algorithm is suitable for optimizing C and kernel function parameter γ of the SVM to obtain a better combination of parameters due to its strong global searching capability.[30] Then, the optimal parameters obtained by the SSA optimization algorithm were used to establish the SVM model. The workflow of the SSA-SVM model is shown in Fig. 1.

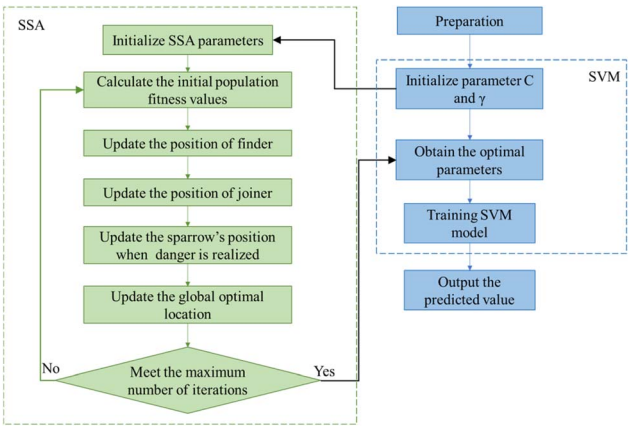

Fig. 1 The working flow chart of the SSA-SVM model.

Performance evaluation. The performance of the models was evaluated in terms of R^2 and the root mean square error (RMSE).[31] Conceptually, a higher R^2 and a lower RMSE indicate a better model accuracy.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{y} - y)^2}{\sum_{i=1}^{N} (\hat{y} - \bar{y})^2} \quad (7)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y} - y)^2}{N}} \quad (8)$$

where $\hat{y}$, y, and $\bar{y}$ are the predicted, actual, and mean values of the target feature, respectively; N is the total number of data points.

Results and discussion

Dataset description

An overview statistical distribution of the dataset was presented in Fig. 2. Table S2† is the brief description of the dataset, including the unit, count, range, mean value and standard deviation of all the features.

As shown in Fig. 2 and Table S2,† the VM content was ranging from 32.2 wt% to 96.4 wt% with an average value of

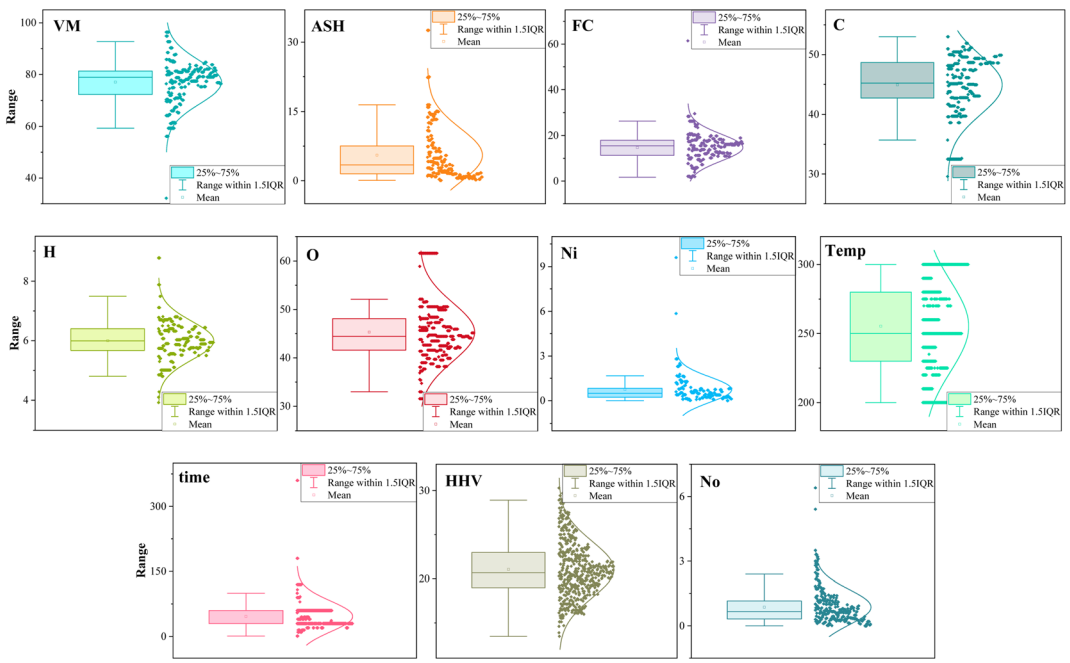

Fig. 2 The statistical distribution of each variable involving the inherent properties of the raw biomass, the torrefaction conditions, and the HHV and No of torrefied biomass.

77.02 wt%. A peak density at 1.76 wt% was observed for ASH content in the range of 0–32.58 wt%. The FC content was distributed between 1.67 wt% and 61.4 wt% with a peak value at 16.51 wt%. The content of O element ranged from 11.37 wt% to 61.55 wt% with a mean value of 43.83 wt%. Such high O content is responsible to the low HHV of the raw biomass. The highest Ni was 14.29 wt%, however, it was less than 0.8 wt% for most of the feedstock employed in the related literatures. As for the torrefaction conditions, the most frequently used. Duration time was 30 min followed by 60 min. Besides, 300 °C was the most favourite temperature.

As for the properties of torrefied biomass, the HHV ranged from 13.48 MJ kg^{-1} to 30.3 MJ kg^{-1} with a median of 20.67 MJ kg^{-1}. No was generally higher than Ni, while it was less than 1.5 wt% for most of torrefied biomass.

Fig. 3 illustrates the Pearson correlation matrix, the detailed information of which was given in Table S3.†

A relatively strong negative correlation was found between ASH and VM (−0.73). Thus, ASH would be removed from the features in the following model. For any two of other features, there was no significant linear correlation between them since all the PCC absolute values were lower than 0.7. It is worth noting that No was strongly linearly correlated to Ni with a positive PCC value of 0.93, indicating that a raw biomass sample with higher nitrogen content can directly result in a higher nitrogen content in their torrefied product.

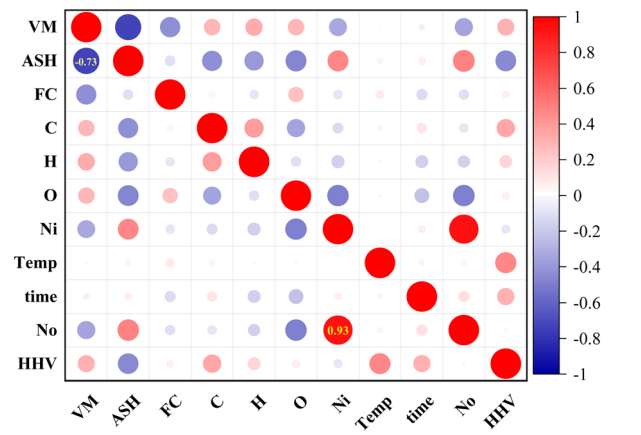

Fig. 3 Pearson correlation matrix between any two features.

Model prediction

Fig. 4 and 5 are the comparison between predicted values and the experimental data for SVM model and SSA-SVM model, respectively. For SVM model (Fig. 4), the data points of both training and testing sets of HHV and No were dispersedly distributed around the black line ($y = x$). While for the SSA-SVM model, all the data points were densely distributed along the $y = x$ line, indicating the equivalence between the predicted values and the experimental data. Therefore, SSA-SVM model had better performance than SVM model in predicting the HHV and No of torrefied biomass, that is, SSA optimization method greatly improved the prediction performance of the SVM model.

The results of R^2 and RMSE for both training set and testing set obtained from SVM and SSA-SVM prediction are shown in Table 1. For both training set and testing set of HHV and No, the R^2 values of SSA-SVM model (>0.91) were larger than the ones of SVM model, indicating that SSA optimization significantly improved the stability and prediction accuracy of the original SVM model. The RMSE values of the two models for both HHV and No were fairly acceptable. While the RMSE values of SSA-SVM model were smaller than the corresponding ones of SVM model, implying a more excellent performance of the SSA-SVM model.

Nieto *et al.*[15,16] predicted the HHV value of torrefied biomass using SVM models. The best values of R^2 and RMSE for SA optimized SVM model was 0.9028 and 0.5171, respectively, and they were 0.9427 and 0.3944 for PSO optimized SVM model. Thus, comparing the results of R^2 and RMSE for HHV predicted using SVM related models in existing literatures with the values obtained in this study, the SSA-SVM model exhibited comparable performance.

While for the prediction of No, to the authors' knowledge, it was performed for the first time in this study.

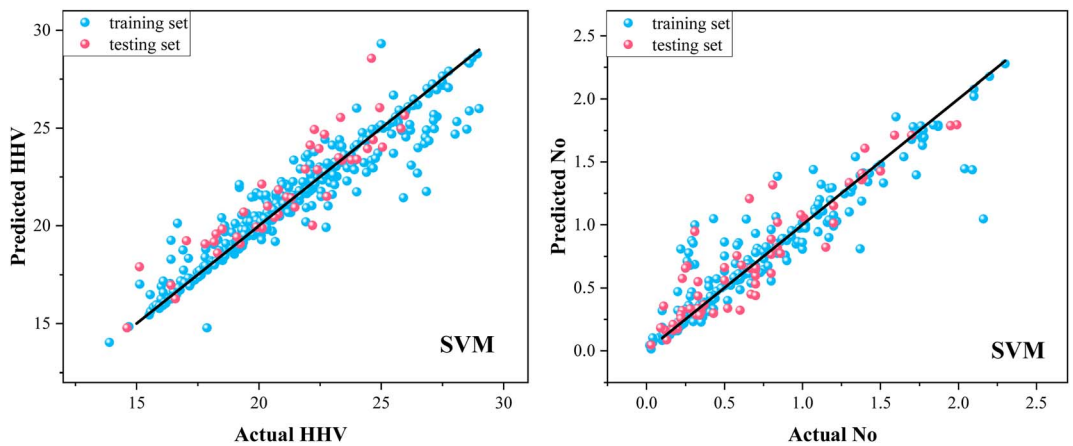

Fig. 4 Comparison of SVM predicted values and experimental data.

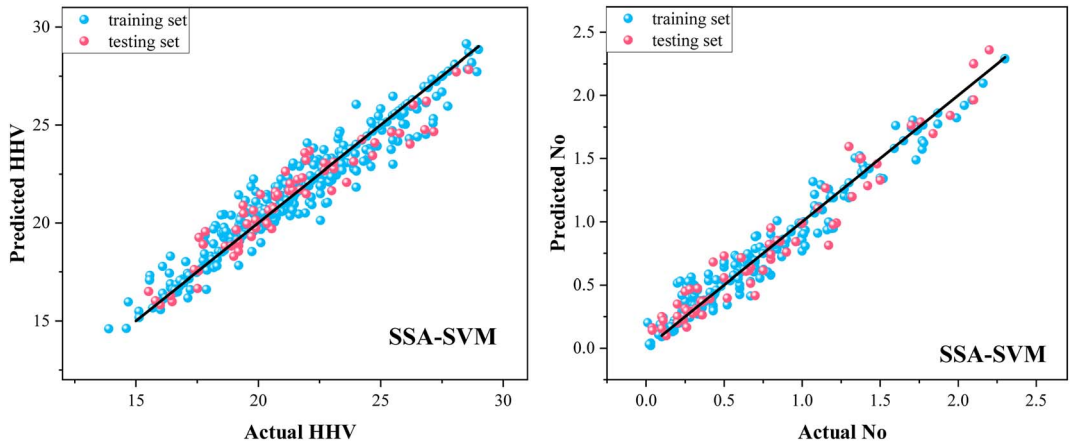

Fig. 5 Comparison of SSA-SVM predicted values and experimental data.

Table 1 R^2 and RMSE results of the two models

	HHV		No	
	SVM	SSA-SVM	SVM	SSA-SVM
R^2 (training)	0.9013	0.9272	0.8838	0.9379
RMSE (training)	4.3346	1.1184	0.395	0.1046
R^2 (testing)	0.8284	0.9111	0.8274	0.918
RMSE (testing)	7.8543	1.2745	0.467	0.16

Fig. 6 illustrates the data points of HHV and No obtained from SVM and SSA-SVM prediction and their comparison with the experimental values. It is obviously that the curve behaviours of SSA-SVM predicted data were more consistent to the actual values for both HHV and No. Therefore, SSA optimization was an efficient method to improve the prediction ability of SVM model.

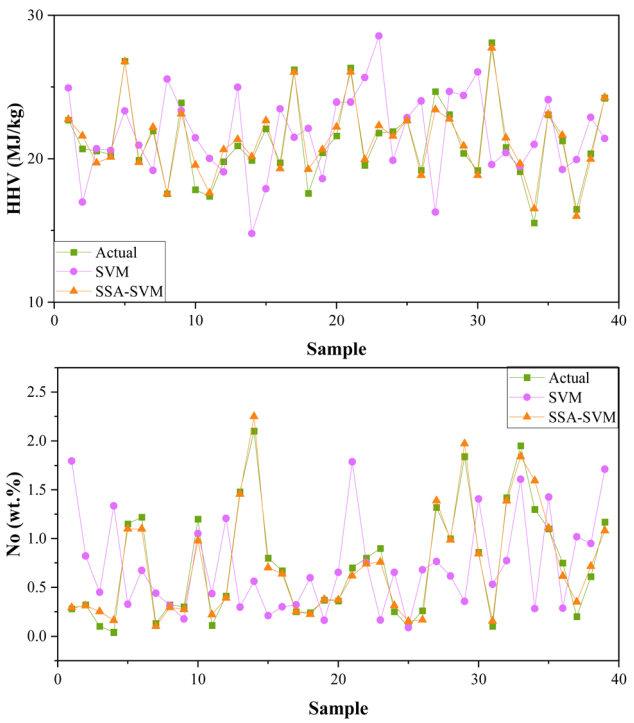

Fig. 6 Comparison between the actual and SVM/SSA-SVM predicted targets of testing set.

Conclusions

In this study, an innovative SVM model with RBF kernel function hybridized with SSA optimization technique was developed to predict the HHV and No values of torrefied biomass for the fuel purpose based on the feedstock properties and the torrefaction conditions. Comparing with the original SVM model, the SSA optimized model exhibited a higher prediction precision for both HHV and No. Higher R^2 values (>0.91) were obtained for SSA-SVM model and the values of MASE were also fairly acceptable, indicating that SSA optimization significantly improved the performance of SVM model. The high predictive precision of the SSA-SVM model was further demonstrated by the agreement between experimental data and predicted values.

Author contributions

Liu Xiaorui: conceptualization, methodology, visualization, writing—original draft preparation, writing—review and editing, funding acquisition. Yang Jiamin: methodology, writing—original draft preparation. Yuan Longji: methodology, validation.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was funded by the Natural Science Foundation of Jiangsu Province, grant number BK20210511.

Notes and references

1 M. W. Seo, S. H. Lee, H. Nam, D. Lee, D. Tokmurzin, S. Wang and Y. Park, *Bioresour. Technol.*, 2022, **343**, 126109.

2 W. Chen, B. Lin, Y. Lin, Y. Chu, A. T. Ubando, P. L. Show, H. C. Ong, J. Chang, S. Ho, A. B. Culaba, A. Pétrissans and M. Pétrissans, *Prog. Energy Combust. Sci.*, 2021, **82**, 100887.
3 Y. Liu, E. Rokni, R. Yang, X. Ren, R. Sun and Y. A. Levendis, *Fuel*, 2021, **285**, 119044.
4 J. González-Arias, X. Gómez, M. González-Castaño, M. E. Sánchez, J. G. Rosas and J. Cara-Jiménez, *Energy*, 2022, **238**, 122022.
5 C. Lokmit, K. Nakason, S. Kuboon, A. Jiratanachotikul and B. Panyapinyopol, *Biomass Convers. Biorefin.*, 2022, DOI: 10.1007/s13399-021-02132-2.
6 Y. Lin, W. Chen, B. Colin, A. Pétrissans, R. Lopes Quirino and M. Pétrissans, *Fuel*, 2022, **310**, 122281.
7 R. Tu, Y. Sun, Y. Wu, X. Fan, S. Cheng, E. Jiang and X. Xu, *Energy*, 2022, **238**, 121969.
8 L. Zhang, Z. Wang, J. Ma, W. Kong, P. Yuan, R. Sun and B. Shen, *Fuel*, 2022, **310**, 122477.
9 F. Kartal and U. Özveren, *Renewable Energy*, 2022, **182**, 578–591.
10 S. Yu, H. Kim, J. Park, Y. Lee, Y. K. Park and C. Ryu, *Int. J. Energy Res.*, 2022, **46**, 8145–8157.
11 H. N. Guo, S. B. Wu, Y. J. Tian, J. Zhang and H. T. Liu, *Bioresour. Technol.*, 2021, **319**, 124114.
12 R. Aniza, W. Chen, F. Yang, A. Pugazhendh and Y. Singh, *Bioresour. Technol.*, 2022, **343**, 126140.
13 H. Y. Ismail, S. Fayyad, M. N. Ahmad, J. J. Leahy, M. Naushad, G. M. Walker, A. B. Albadarin and W. Kwapinski, *J. Cleaner Prod.*, 2021, **326**, 129020.
14 F. Kartal and U. Özveren, *Biomass Bioenergy*, 2022, **159**, 106383.
15 P. J. García Nieto, E. García-Gonzalo, J. P. Paredes-Sánchez, A. Bernardo Sánchez and M. Menéndez Fernández, *Neural. Comput. Appl.*, 2019, **31**, 8823–8836.
16 P. J. García Nieto, E. García Gonzalo, F. Sánchez Lasheras, J. P. Paredes Sánchez and P. Riesgo Fernández, *J. Comput. Appl. Math.*, 2019, **357**, 284–301.
17 T. Onsree and N. Tippayawong, *Renewable Energy*, 2021, **167**, 425–432.
18 T. Onsree, N. Tippayawong, S. Phithakkitnukoon and J. Lauterbach, *Energy*, 2022, **249**, 123676.
19 E. Leng, B. He, J. Chen, G. Liao, Y. Ma, F. Zhang, S. Liu and J. E, *Energy*, 2021, **236**, 121401.
20 T. Williams, K. McCullough and J. A. Lauterbach, *Chem. Mater.*, 2020, **32**, 157–165.
21 L. Xiaorui, Y. Xudong, X. Guilin and Y. Yiming, *Fuel*, 2021, **291**, 120264.
22 X. Liu, Z. Luo, C. Yu and G. Xie, *Fuel*, 2019, **246**, 42–50.
23 J. Li, X. Zhu, Y. Li, Y. W. Tong, Y. S. Ok and X. Wang, *J. Cleaner Prod.*, 2021, **278**, 123928.
24 Q. Tang, Y. Chen, H. Yang, M. Liu, H. Xiao, S. Wang, H. Chen and S. Raza Naqvi, *Bioresour. Technol.*, 2021, **339**, 125581.
25 Z. Ullah, M. Khan, S. Raza Naqvi, W. Farooq, H. Yang, S. Wang and D. N. Vo, *Bioresour. Technol.*, 2021, **335**, 125292.
26 M. S. Nick Guenther, *Stata J.*, 2016, **4**, 917–937.
27 Y. Wang, Z. Liao, S. Mathieu, F. Bin and X. Tu, *J. Hazard. Mater.*, 2021, **404**, 123965.
28 C. Yin, X. Deng, Z. Yu, Z. Liu, H. Zhong, R. Chen, G. Cai, Q. Zheng, X. Liu, J. Zhong, P. Ma, W. He, K. Lin, Q. Li and A. Wu, *Biotechnol. Biofuels*, 2021, **14**, 106.
29 J. Xue and B. Shen, *Syst. Sci. Control. Eng.*, 2020, **8**, 22–34.
30 W. Tuerxun, X. Chang, G. Hongyu, J. Zhijie and Z. Huajian, *IEEE Access*, 2021, **9**, 69307–69315.
31 X. Yuan, M. Suvarna, S. Low, P. D. Dissanayake, K. B. Lee, J. Li, X. Wang and Y. S. Ok, *Environ. Sci. Technol.*, 2021, **55**(17), 11925–11936.