Environmental Technology & Innovation 37 (2025) 104012

Contents lists available at ScienceDirect

Environmental Technology & Innovation

journal homepage: www.elsevier.com/locate/eti

Biochar energy prediction from different biomass feedstocks for clean energy generation

Nikhil Pachauri [a], Chang Wook Ahn [b], Tae Jong Choi [c,*]

[a] Department of Mechatronics, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[b] Artificial Intelligence Graduate School, Gwangju Institute of Science and Technology, Gwangju 61005, Republic of Korea
[c] Graduate School of Data Science, Chonnam National University, Gwangju 61186, Republic of Korea

ARTICLE INFO

Keywords:
Higher heating value
SVR
GPR
Biomass: voting ensemble
Optimization

ABSTRACT

This paper presents a novel approach for predicting various feedstock higher heating values (HHV) using a voting ensemble machine-learning model. The proposed model, referred to as VSGB, combines Support Vector Regression (SR), Gaussian Process Regression (GR), and Boosting (BO) using a weighted sum technique. The Invasive Weed Optimization (IWO) algorithm is employed to estimate hyperparameter values of the VSGB model. Moreover, comparative performance analysis is conducted using several models, such as linear regression (LR), generalized additive model (GAM), bagging (BAG), decision tree (DT), and neural network (NN). The simulation findings demonstrate that the VSGB has a high level of accuracy in predicting the HHV derived from biomass waste. This is evidenced by the lower Root Mean Square Error (RMSE) and Average Absolute Relative Difference (AARD%) values (0.813 and 2.827%, respectively) compared to other Machine Learning (ML) predictive models. Additionally, the present study establishes an empirical correlation between the higher heating value (HHV) and the input characteristics carbon (C), hydrogen (H), oxygen (O), nitrogen (N), and sulphur (S) through the utilization of the IWO algorithm.

1. Introduction

There has been a growing interest in exploring environmentally sustainable alternatives to traditional methods of energy generation, driven by concerns over the impending energy crisis and the detrimental impact of fossil fuel combustion on the environment (Vargas-Moreno et al., 2012; Adeniyi et al., 2019). The primary advantages of utilizing biomass as an energy source include its sustainability, easy accessibility, and cost-effectiveness (Zhu et al., 2024; Fei et al., 2023; Chen et al., 2023; Guo et al., 2023; Guo et al., 2023). Simultaneously, the issue of biomass disposal will be simplified (Demiral, 2009; Adeniyi and Ighalo, 2020). Biofuels obtained from biomass release CO_2 after combustion, but carbon in biochar remains sequestered due to stabilization via pyrolysis. Moreover, biochar substantially enhances soil health by augmenting water retention, nutrient accessibility, and microbial activity. It will help to diminish the dependence on chemical fertilizers and foster agricultural sustainability. Biochar also reduces methane and nitrous oxide emissions from soils, which are potent greenhouse gases. Biochar production is versatile, as feedstock is obtained from many agricultural and forestry residues. This enhances waste management. Biochar formation reduces greenhouse gas emissions, especially when employing a closed-loop system, making it a more sustainable option than biofuel production. Recent reports from the

* Corresponding author.
E-mail addresses: nikhil.pachauri@manipal.edu (N. Pachauri), cwan@gist.ac.kr (C.W. Ahn), ctj17@jnu.ac.kr (T.J. Choi).

https://doi.org/10.1016/j.eti.2024.104012
Received 7 June 2024; Received in revised form 17 November 2024; Accepted 30 December 2024
Available online 2 January 2025

International Biochar Initiative (IBI) [biochar-international.org/fy24-impact-snapshot]. predict global biochar production of approximately 350,000 metric tons yearly, estimated at 0.5–1 million tons. It indicates substantial prospects for development in the forthcoming years. Further, a report published by the International Renewable Energy Agency (IRENA) states that biochar is utilized to produce briquettes as a substitute for wood coal. Mandulis Energy has commenced 8 MW (16 locations of 500 kW each) of off-grid electricity and manufacturing of briquettes [www.irena.org/media/Files/IRENA/Agency/Publication/2019/Jun/IRENA_G20_climate_sustainability_2019.pdf]. It is crucial to know the heating value of the biomass under consideration before its utilization in any thermochemical energy system. The higher heating value is commonly referred to as HHV. It is a metric utilized to assess the overall energy content of fuel (Ahmaruzzaman, 2008; Channiwala and Parikh, 2002). Furthermore, HHV is essential in optimizing biomass conversion processes across various conditions. Different tools based on mathematical frameworks, simulations, and optimization are developed to analyze and advance intricate biomass conversion technology (Puig-Arnavat et al., 2010). It is imperative to ensure the availability of methods that facilitate the assessment of biomass potential energy and determining decentralization product formulations for apprehensive fuels without empirical data (Wen et al., 2017). This can be achieved by experimenting with a bomb calorimeter, which is time-consuming and costly compared to mathematical and simulation-based methods. Therefore, it is necessary to develop cost-effective and easy-to-implement strategies for designing biomass energy systems (Xing et al., 2019). The progress made in artificial intelligence, machine learning, and deep learning has enabled engineers to create predictive models that establish a correlation between higher heating value (HHV) and input data derived from both ultimate and proximate analysis. Table 1 shows the literature surveyed related to the algorithms used to predict the biochar of different biomass.

(Zhu and Yang, 2022) have proposed ANN and PSO-optimized ML models to forecast municipal solid waste's heating values. The ANN model that integrates both proximate and ultimate analyses of municipal solid waste (ANN-4) achieved the highest accuracy, with an error rate below 10 %. (Timilsina et al., 2024) employ a machine learning model to determine the HHV as a critical predictive tool for assessing waste-to-energy systems. Six models were employed, demonstrating exceptional predictive accuracy with R^2 values between 0.83 and 0.98. The analysis indicates that carbon and hydrogen enhance HHV, but oxygen and ash content diminish it, providing comprehensive advice for decision-making regarding waste-to-energy utilization. (Kocer, 2024), proposed a multilinear regression algorithm to predict HHV for proximate and ultimate analysis. The highest value of R^2 achieved is 0.997 for the regression model designed for ultimate analysis. (Liu et al., 2024), use random forest (RF) and extreme gradient boosting (XGboost) algorithms for bio-oil product yield and HHV prediction. Results suggested that the highest values of R^2 are 0.942 and 0.940 for XGboost. (Brandić et al., 2024) suggested an artificial neural network model with a 2–12–1 architecture for predicting HHV, utilizing a dataset derived from biomass's ultimate and structural analysis. The proposed ANN structure has attained a superior R^2 value of 0.96 compared to previous ML models. (Wang et al., 2024), use an XGBoost and ANN for the HHV prediction based on the characteristics of biomass and pyrolysis conditions. Results show that XGboost attains $R^2 = 0.83–0.94$. (Kandpal et al., 2024) employed three ensemble learning techniques (RF, XGboost, and Adaptive Boosting (AdaBoost)) to construct models for predicting biochar production and HHV. The RF model achieved a satisfactory R^2 of 0.86, whereas XGB produced the optimal model with an R^2 of 0.87 for HHV. SHAP analysis identified pyrolysis temperature and ash content as the two most significant features. A Bayesian-optimized Gaussian process regressor (GPR) was developed by Kaya et al. (2024), to predict the HHV based on biomass characteristics and process conditions. It achieves the lowest value of MAE (0.4435) compared to other designed ML models. (Ighalo et al., 2024) conducted a mini-review of synthesized research on biomass HHV prediction using MLP-ANN models, highlighting their accuracy, data dependence, and future potential.

1.1. *Significant contribution and novelty*

There is an increasing interest in discovering eco-friendly alternatives to conventional methods of energy generation, driven by

Table 1
A literature survey was conducted for an algorithm to predict biomass's biochar.

Authors	Algorithms	Results claimed
(Onsree et al., 2022)	Gradient boosting tree algorithm (GBT)	The root mean square (RMSE) attained by GBT is 0.79 for HHV prediction compared to other ML algorithms.
(Güleç et al., 2022)	Different ANN models based on Levenberg and Bayesian learning algorithm	Results suggested that the ANN models designed for combined ultimate and proximate analysis data sets attained a higher value of R^2 (0.962 and 0.876) compared to other cases.
(Mu et al., 2022)	Particle swarm optimized neural network (PS-NN)	PS- NN has a higher value of R^2 (0.85) w.r.t. artificial neural network (ANN).
(Ighalo et al., 2020)	Linear regression (LR)	The RMSE achieved by the LR is 8.15, which is lower than other applied ML.
(Jakšić et al., 2023)	ANN based on different learning algorithms	Levenberg–Marquardt predicts the HHV with the lowest value of MSE of 1.17.
(Brandić et al., 2022)	Ten different ANN models are developed.	Results revealed that Model 9 is the best among all the designed ANN models, with the highest R^2 of 0.47.
(Dai et al., 2021)	Extreme machine learning (EML)	EML achieved the highest value of R^2 (0.989) compared to ANN.
(Dubey and Guruviah, 2023)	Random forest (RF)	The result shows that the prediction accuracy of RF is better than other ML in terms of RMSE (1.4204).
(Kartal and Özveren, 2022)	ANN is designed for energy prediction	Results show that ANN achieved 4 % less value of MAE.
(Katongtung et al., 2022)	Extreme gradient boosting (XGboost)	XGboost has a higher range of R^2 (0.87–0.90) compared to other ML algorithms.

concerns about the imminent energy crisis and the detrimental impact of burning fossil fuels on the environment. Biomass is an upcoming energy source due to its sustainability, accessibility, and cost-effectiveness. Furthermore, the process of managing biomass disposal becomes more straightforward. Before utilizing biomass in any thermochemical energy system, it is crucial to comprehend its heating value, commonly known as HHV. This statistic facilitates the assessment of the fuel's energy content. Moreover, HHV is pivotal in enhancing biomass conversion processes under diverse situations. Researchers are now developing machine learning-based tools to study and improve complex biomass conversion technologies.

Research Gap: - The literature review indicates that multiple ML techniques have been employed to develop a predictive model for HHV estimation based on biochar's ultimate and proximate analysis. However, every ML algorithm has pros and cons, as the predicting performance of the ANN depends on various factors, including learning rate, number of hidden layers, and neurons in each hidden layer. The XGBoost algorithm encounters challenges when applied to sparse and unstructured datasets. The performance of the techniques varies depending on the parametric and non-parametric applications. The choice of the number of trees and leaf nodes in each tree can influence the RF performance. This implies that the ideal selection of its hyperparameter values and applications significantly impacts every machine learning algorithm's performance. Therefore, selecting an ML algorithm should be examined on a case-by-case basis (Pachauri and Ahn, 2023). The scalability of the ML models is critical in practical applications; machine learning models must efficiently manage varying operating circumstances and datasets from diverse feedstocks. Moreover, most models are optimized for a singular feedstock type. A model trained on diverse feedstock types is required. A single ML algorithm will only work efficiently in some applications. Therefore, this work designs and implements a voting ensemble model for better scalability and generalization.

Novelty and contribution: In this article, a voting ensemble ML algorithm-based predictive model is proposed to predict the HHV of biomass. The voting ensemble technique is designed by combining Gaussian process regression (GR), support vector regression (SR), and boosting (BO) algorithms, leading to VSGB. A voting ensemble will predict the HHV by taking the weighted average of the individual predictions of GR, SR, and BO. Furthermore, the hyperparameters of GR (sigma (σ)), SR (Box Constraint, Kernel Scale, Epsilon), BO (learning rate (η)), and three weighting functions (*w1, w2,* and *w3*) play a vital role in the accurate prediction of HHV. Estimation of the optimal values of these parameters is a complex task. Therefore, an optimization technique known as invasive weed optimization (IWO) (Mehrabian and Lucas, 2006) is utilized to estimate the optimal values of σ, Box Constraint, Kernel Scale, Epsilon, η, *w1, w2*, and *w3*, respectively. The IWO algorithm is more likely to avoid local minima positions compared to GA and PSO because the IWO methodology has a continuous dispersion architecture and is normally distributed across the search space. Additionally, the IWO algorithm contains a declining variance parameter that centers on each parent plant (Barisal and Prusty, 2015). The subsequent points encapsulate the principal contribution of the work.

- A voting ensemble model, which is an amalgamation of GR, SR, and BO leads to VSGB, is recommended for the biochar HHV prediction of biomass. IWO calculates the optimum values Box Constraint, σ, Epsilon, η, Kernel Scale, *w1*, *w2*, and *w3*, respectively.
- A comprehensive comparative analysis was conducted to assess the predictive accuracy of the VSGB model in comparison to LR, GAM, BAG, DT, and NN. Further, the performance of VSGB is compared with that of ML models in the literature.
- Finally, the empirical relationship between the ultimate analysis components and HHV is estimated using IWO. The result is then compared with other empirical relationships specified in the literature.

The rest of the paper is organized as follows: method and material description are given in Section 2, results and discussion are discussed in Section 3, and Section 4 presents the article's conclusion.

2. Method and material

In this section, a description of the dataset utilized for this work, along with the voting ensemble technique, will be discussed. Furthermore, the IWO optimization algorithm will also be discussed briefly.

2.1. Biochar dataset

The dataset employed in this study comprises five input parameters, namely carbon (C), hydrogen (H), oxygen (O), nitrogen (N), and sulfur (S), along with one output parameter, namely HHV (See the supplementary file, Table S1, for further details). A combined collection of specimens was grouped, with eighty-two specimens falling under the category of fruit by-products, one hundred eighty samples categorized as agriculture waste, one hundred sixty-eight specimens originating from wood or tree species, fifty-one specimens obtained from leaves or fibrous material, and thirty-seven specimens classified as other forms of biomass waste such as animal waste, wastewater sludge, and aerobic digestion, among others. Thirty-three specimens are categorized as briquettes, charcoal, and pellets, nineteen as cereals, and seventeen as industrial waste. Furthermore, Agricultural farms involved in several categories of biomass products represent a spectrum of activity. Each generates distinct categories of waste. Citrus, banana, and mango plantations generate by-products, including peels and seeds. Agricultural operations cultivating staple crops such as rice, wheat, and maize produce agricultural byproducts post-harvest, including straw and husks. Wood waste comprises branches, bark, and sawdust from timber -forestry operations. Fiber crops, like cotton, jute, and hemp, generate fibrous material waste. Livestock farms that rear cattle, chickens, or pigs produce animal waste from manure and mixed bedding materials. Some farms with wastewater treatment or anaerobic digestion systems, particularly dairy and pig farms employing waste-to-energy systems, contain sludge and digestion residue. Facilities that transform agricultural or wood waste into solid fuels produce briquettes, charcoal, and pellets, utilizing rice husks

or coconut shells. Conventional cereal crops, such as wheat, barley, and maize, provide by-products, including husks and straw. Diverse categories of farm waste are transformed into distinct biomass types, each possessing unique features, facilitating energy and material recovery for sustainable agriculture and efficient waste management. Table 2 shows the statistical information for the given dataset. Fig. 1 shows the histogram representation of the dataset

Fig. 2 shows the Pearson Correlation analysis between input and output attributes. A positive value of correlation shows the direct relationship between input and output. However, negative values describe the inverse relationship. It can be observed that the C (0.9439) and S (0.08304) have a positive correlation with HHV. On the other hand, O (-0.5504), H (-0.2173), and N (-0.1319) have a negative correlation. It is also observed that the correlation between input attributes is very low, which means there is no multi-collinearity in the dataset. In addition, the correlation of attributes with HHV allows for calculating each feature's contribution to final predictions, considerably improving interpretability. This is particularly beneficial in fields such as energy modeling, where comprehending the factors influencing energy content is essential.

2.2. Support vector regression

Support Vector Regression (SR) is a supervised machine learning algorithm that utilizes SVM's principle to forecast distinct values. The fundamental principle underlying the concept of SR is to determine the optimal linear regression line (Taki and Rohani, 2022). The hyperplane, also known as the best-fit line, is characterized by having the maximum number of data points. The concept of SR generally encompasses two distinct categories of hyperplanes: positive and negative. The hyperplane is referred to as positive when it is located on the decision border's positive side and negative when positioned on the negative side. Moreover, the boundaries of the hyperplanes correspond to the support vectors, which indicate the data points near the hyperplane. Estimating the direction and location of the hyperplane is beneficial. The mathematical representation of SR function is expressed as follows (Cao et al., 2016).

$$f(x) = \quad (c, \varsigma(x)) + d \tag{1}$$

$\varsigma(x)$Nonlinear term
cweight vector
dbias vector
The optimization equation for SR can be written as follows.

$$\left\{ \begin{array}{lll} & \min\limits_{c,u,e} f(c,e) = \frac{1}{2}\|c\|^2 + \frac{1}{2}\gamma \sum\limits_{i=1}^{M} e_i^2 & \\ s.t. z_k & = \quad \langle c, \varsigma(x_i) \rangle + u + e_i \quad i = 1, 2, 3, \ldots M & \end{array} \right\} \tag{2}$$

where γ is a regulation factor, and e_i is an error for the M training dataset. Moreover, the Lagrangian method is utilized to approximate the optimization problem.

$$L(c, u, e, \alpha) = \frac{1}{2}\|c\|^2 + \frac{1}{2}\gamma \sum_{i=1}^{M} e_i^2 - \sum_{i=1}^{M} \alpha_i \{ \langle c, \varsigma(x_i) \rangle + u + e_i - z_i \quad \} \tag{3}$$

where α_i is the Lagrange multiplier. A detailed description of SR is given in (Cao et al., 2016).

2.3. Gaussian process regression

GR is a collection of fixed quantities characterized by random traits following an ensemble Gaussian distribution. The primary objective of GR is to utilize the existing data to ascertain a substantial correlation between input and output attributes, thereby fulfilling the following equation (Schulz et al., 2018).

$$a = g(b) + \beta \tag{4}$$

where a and b are the output and input attributes, $\beta \epsilon \left(M\left(0, \rho_\beta^2\right) \right)$, it is supposed that $g(b)$ is spread as a Gaussian distribution (GD). The GD description is facilitated by the utilization of a mean and covariance as follows: -

Table 2
Statistical Information of the Dataset.

Variables	Average	Kurtosis	Standard deviations
C	48.205	15	5.909
H	5.468	7.806	0.744
O	38.648	11.016	7.119
N	1.075	9.829	1.170
S	0.232	6.419	0.193
HHV	19.159	13.571	2.524

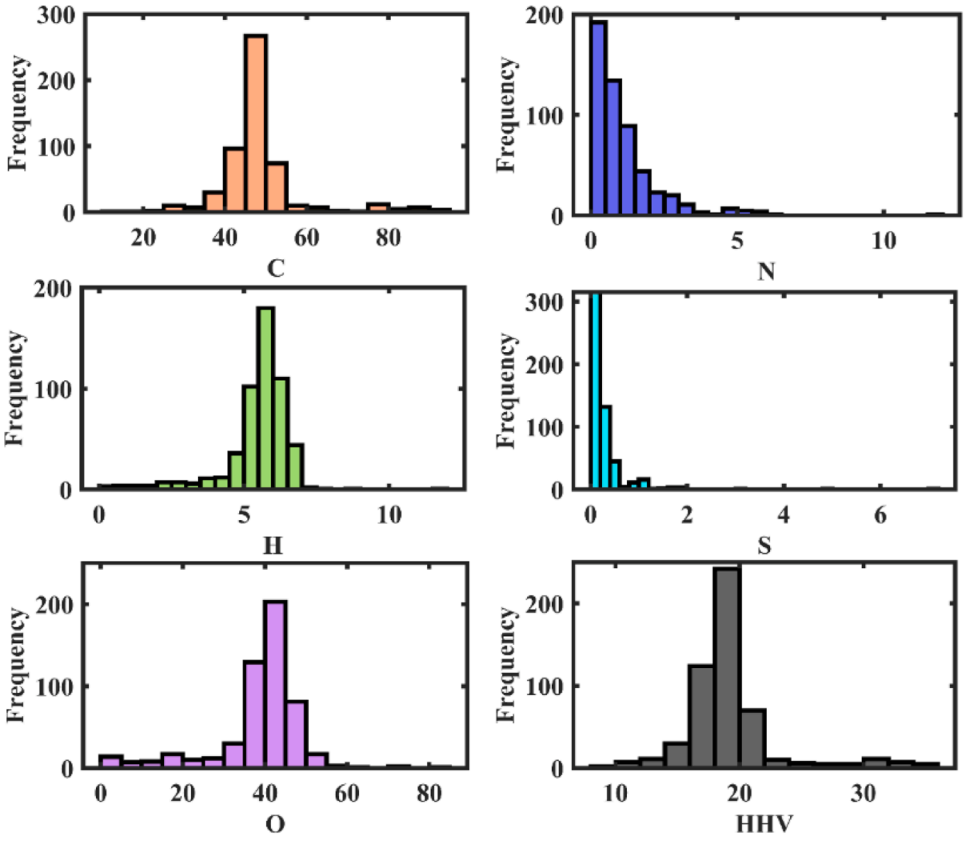


Fig. 1. Histogram representation of the dataset.

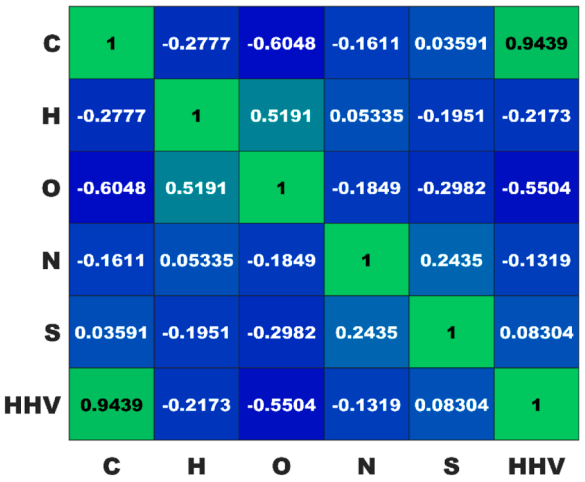


Fig. 2. Pearson Correlation analysis between input and output attributes.

$$g(b) = GD(n(b), l(b, b')) \tag{5}$$

The average function, denoted as *n(b)*, represents the mean value of all the functions within the given variations at the input *b* (Schulz et al., 2018).

$$mean(b) = \quad F[g(b)] \tag{6}$$

The covariance *l(b, b')* analyzes the interdependence of function values across input attributes *b* and *b'*.

$$cova(b,b') = \quad F[(g(a)-n(b))(g(b')-l(b'))] \tag{7}$$

where l is the kernel of the GR. A total of 5 kernels may be used for GR *i.e.,* rational quadratic, squared exponential, exponential, matern 3/2, and matern 5/2, respectively.

2.4. Boosting

The BO technique is an ensemble approach that aims to minimize biases rather than variance sequentially (Pachauri et al., 2022). The boosting technique involves constructing a forecasting model using the primary dataset and then iteratively enhancing the algorithm to reduce the variance of previous models. BO employs multiple models characterized by low variance and substantial biases to enhance the accuracy of forecasts. It enhances the predictive capabilities of a weak learner. The BO used a tree-based technique to construct a robust regression model using a gradient-boosting strategy. BO output is achieved by combining forecasts generated by multiple models. The models are developed in a sequence, where each model is trained to minimize the residual error by incorporating the predictions of all preceding models (Pachauri and Ahn, 2023). Furthermore, a least square gradient-based BO algorithm is used in this work. The mathematical formulation of the algorithm is as follows.

Step 1. Least square boosting will reduce the error between the predicted $y_i^{'}$ and actual value of the output y_i for n number of samples.

$$LOSS = \sum_{i=1}^{n}\left(y_i - y_i^{'}\right)^2 \tag{8}$$

Step 2. The algorithm starts with a simple model with constant prediction.

$$\left(y_i^{'}\right)^{(0)} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{9}$$

$\left(y_i^{'}\right)^{(0)}$ is the predicted value of each datapoint at i for 0^{th} model.

Step 3. In iteration m, it builds a new model $f_m(x)$ to fit the error of the previous model and add it to the ensemble of models with the learning rate α

- Calculate the error

$$(r_i)^{(m)} = \left(y_i - \left(y_i^{'}\right)^{(m-1)}\right) \tag{10}$$

- Fit a weak learner

$$f_m(x) \approx (r_i)^{(m)} \tag{11}$$

- Update the model

$$\left(y_i^{'}\right)^{(m)} = \left(y_i^{'}\right)^{(m-1)} + \alpha f_m(x_i) \tag{12}$$

Step 4. After the M iteration, the final output is the sum of all the weak learners $f_m(x)$ for a given input x_i

$$y_i^{'} = \sum_{m=0}^{M} \alpha f_m(x_i) \tag{13}$$

2.5. Voting ensemble

The voting ensemble is an ML technique combining predictions from multiple methods. This approach is commonly utilized to enhance the system's efficacy by incorporating multiple models rather than relying solely on one learning method. In essence, majority voting is employed in classification and regression problems, which involves aggregating the selections made by various ML approaches (Chen et al., 2022). The fundamental idea of voting regression is to take the average of all predictions made using different ML techniques. The disadvantage of this strategy is that it will reduce total prediction accuracy if one ML technique predicts with a higher error. Therefore, this work considers a weighted average of all the techniques' predictions a final prediction. In this method, the weightage of the ML with higher error is less in the final prediction than in the other methods, thereby improving the overall prediction.

2.6. Invasive weed optimization

The IWO is a population-based technique that aims to identify the global optimum by simulating the interoperability and unpredictability observed in the colonies of weeds. They are potent botanical species whose aggressive growth patterns significantly threaten crops. They have demonstrated high resistance and adaptability in response to environmental changes. Hence, a robust

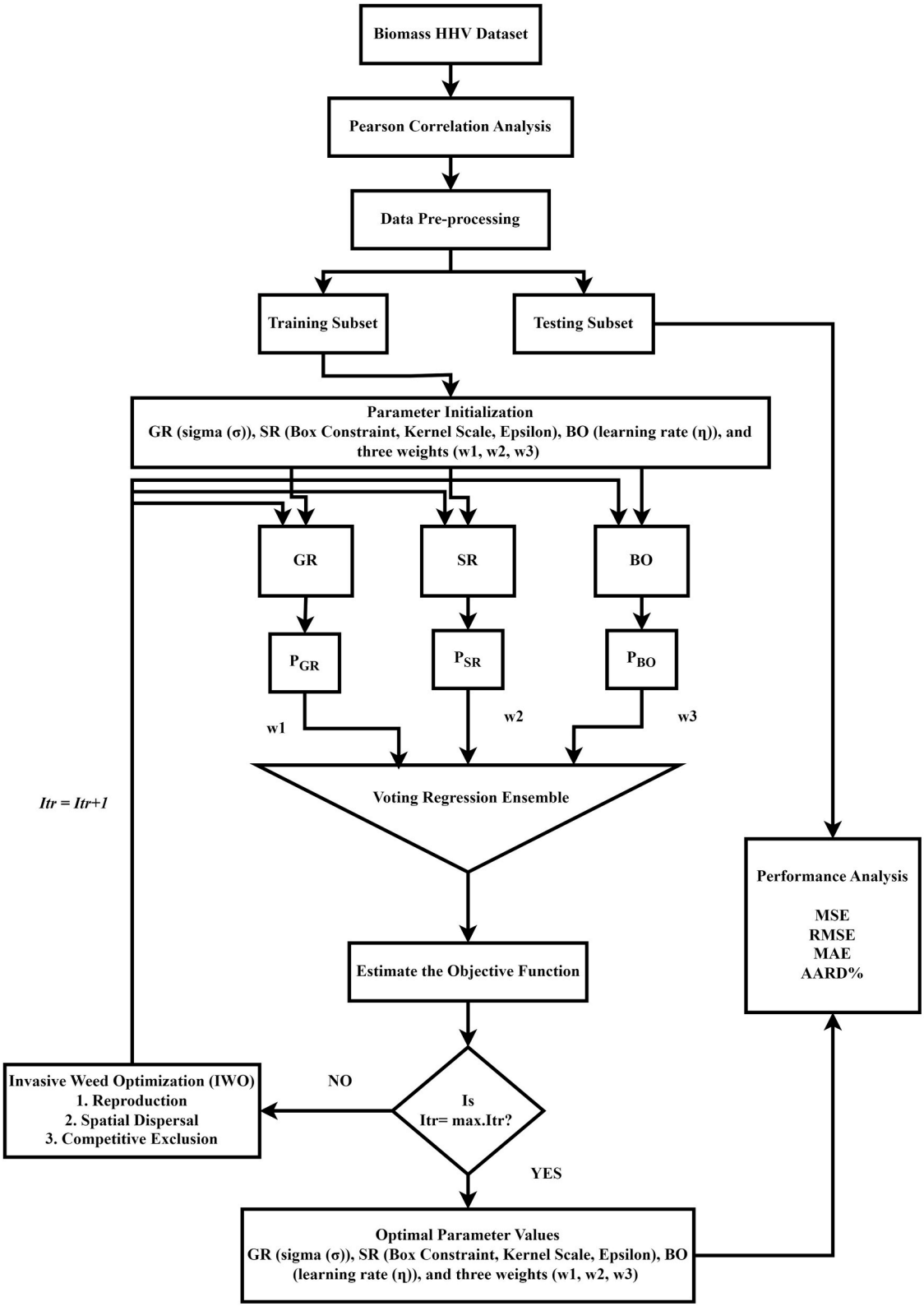


Fig. 3. Flowchart for the IWO optimizes VSGB predictive model.

optimization algorithm is derived by considering their distinctive attributes. The algorithm developed aims to replicate the characteristics of resistance, versatility, and unpredictability observed within a weed community specimen (Mehrabian and Lucas, 2006). The following steps are applied to simulate the colonizing behavior of invasive weeds.

1. The first step is to initialize the population in each search space with dimensionality *w*.
2. The reproductive capacity of an individual plant within a population is determined by its health and the whole colony's collective health. This is achieved through a gradual rise in seed generation, ranging from a minimum to the maximum number of seeds each plant produces. In essence, a plant's reproductive output is determined by its level of adaptability, as well as the range of adaptability within the colony, to maintain a linear growth pattern. The inclusion of this step introduces a notable characteristic of the optimization technique. In applying optimization for a specific problem, it is commonly perceived that possible individuals, instead of infeasible individuals, possess superior fitness values (where "superior" denotes a higher likelihood of survival and reproduction). Consequently, the reproduction of infeasible individuals is typically prohibited.
3. This step, called spatial dispersal, encompasses the concepts of unpredictability and adjustment within the algorithm. The produced seeds are spread arbitrarily across the search space, which has d dimensions. This distribution is achieved using arbitrary numbers that follow a normal distribution, with a mean of zero and varying variances. This implies that seeds will be dispersed randomly near the parent vegetation. Furthermore, as the step of the algorithm proceeds, the value of the *std.* will be decreasing from its initial ($\mu^{\min}$) to the final value (μ^{final}).

$$\mu^{iter} = \frac{\left(itr^{\max} - itr^{\min}\right)^{m}}{\left(itr^{\max}\right)^{m}} \left(\mu^{\min} - \mu^{final}\right) + \mu^{final} \tag{14}$$

where *i*tr . is the no. of iterations, μ^{iter} is the standard deviation, and m is the nonlinear index value.
4. The next step is competitive elimination, in which the quantity of grasses surpasses the predetermined maximum threshold (P_{max}) within the colony, and the grass exhibiting the lowest fitness level is eliminated, thereby ensuring a consistent population of herbs within the colony.
5. This optimization process will run until the stopping criteria of a maximum number of iterations will not be obtained.

Fig. 3 shows the detailed flowchart of the IWO-optimized voting ensemble (VSGB) ML technique for the biochar prediction. The preliminary phase in developing any ML technique involves the preprocessing of the dataset. The procedure begins with acquiring dataset samples from several sources (see supplementary material Table S1). Afterwards, missing values are addressed through either removal or interpolation; nonetheless, the dataset contains no missing entries. Consequently, the dataset guarantees data accuracy by removing duplicates, correcting outliers, and fixing inaccuracies. Ultimately, divide the dataset into training and testing subsets, utilizing cross-validation to evaluate the model's efficacy. According to existing literature, there's no universally applicable guideline for determining the appropriate ranges for data splitting. Multiple authors have examined the diverse spectrums of data., such as 80–20 % (Parikh et al., 2005), 75–25 % (Xing et al., 2019), 83–17 % (Shi et al., 2016), etc., for HHV prediction. Hence, this study partitions the dataset into 80 % for training and cross-validation while allocating the remaining 20 % for testing the proposed ML technique. Furthermore, the hyperparameter of GR (sigma (σ)), SR (Box Constraint, Kernel Scale, Epsilon), BO (learning rate (η)), and three weighting functions (w1, w2, w3) are initialized as per the lower and upper ranges (See supplement material Table S2). The training set is utilized to train GR, SR, and BO in the next step. The 5-fold cross-validation method is commonly employed during training. Subsequently, a voting ensemble is implemented by assigning weights to the three outputs obtained from GR (P_{GR}), SR (P_{SR}), and BO (P_{BO}), respectively. The voting ensemble refers to the process of combining multiple machine learning algorithms through a weighted aggregation mechanism, which can be defined as follows:

$$P_{final} = \frac{(w_1 \times P_{GR} + w_2 \times P_{SR} + w_3 \times P_{BO})}{(w_1 + w_2 + w_3)} \tag{15}$$

The successive step involves the computation of the objective function, specifically the AARD % expressed as a percentage, quantifying the disparity between the predicted and actual values. The process is iteratively executed until the stopping condition of the maximum number of iterations (***Itr.*** = ***max. Itr***) is met. The maximum iteration value (***max. Itr***) is taken as 100. The optimized values for the parameters σ, Box Constraint, Kernel Scale, Epsilon, *η, w1, w2*, and *w3* are determined upon completion of the final iteration. Finally, the performance of the VSGB model will be assessed and compared to other machine learning models by utilizing metrics such as Mean Squared Error (MSE), RMSE, Mean Absolute Error (MAE), and AARD% for testing subsets.

$$MSE = \frac{1}{M}\sum_{j=1}^{M}\left(Y_{j,predicted} - Y_{j,actual}\right)^2 \tag{16}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(Y_{j,predicted} - Y_{j,actual}\right)^2} \tag{17}$$

$$MAE = \frac{1}{M}\sum_{j=1}^{M}\left|\left(Y_{j,predicted} - Y_{j,actual}\right)\right| \tag{18}$$

$$AARD\% = \frac{1}{M}\sum_{j=1}^{M}\frac{\left|\left(Y_{j,predicted} - Y_{j,actual}\right)\right|}{Y_{j,actual}} \times 100 \tag{19}$$

3. Result and discussion

This study offers a voting ensemble (VSGB) for HHV prediction, the weighted average of SR, GR, and BO forecasts. IWO optimization is utilized to estimate the design parameters, σ, Box Constraint, Kernel Scale, Epsilon, η, $w1$, $w2$, and $w3$, of VSGB, respectively. Furthermore, the prediction accuracy comparison of VSGB with LR, GAM, BAG, DT, and NN regarding the MSE, RMSE, MAE, and AARD% values performance indices. In addition to this, IWO is also utilized to calculate the empirical relationship between input and output attributes. Fig. 4(a) and (b) show the convergence graphs obtained while optimizing the hyperparameters of VSGB (training process) and polynomial coefficients of the empirical equation. From Fig. 4(a) and (b), it is feasible to track the progression of the optimization algorithm and ascertain its convergence with the optimal solution. Moreover, it indicates the direction in which the objective function is trending in terms of minimizing its value. The data can be utilized to implement modifications and enhance the optimization process. The optimised parametric values of $\sigma = 0.0001$, Box Constraint = 67, Kernel Scale = 52, Epsilon = 0.1765, $\eta = 0.3276$, $w1 = 0.9984$, $w2 = 0.9388$, and $w3 = 0.4301$, respectively. Simulation studies were performed using MATLAB 2023 on an 11th Gen Intel(R) Core (TM) i5–1135G7 processor operating at 2.40 GHz.

3.1. VSGB performance for HHV prediction

The performance of the VSGB algorithm is evaluated through a comparative analysis with several other single machine learning models, namely LR, GAM, BAG, DT, and NN. The scatter plots of all the designed ML techniques for the testing dataset are depicted in Fig. 5(a). This will establish a correlation between the anticipated and desired values of HHV. The figure demonstrates that the HHV values predicted by VSGB exhibit higher proximity to the target HHV. Furthermore, the correlation coefficient (R) value achieved by VSGB (0.946) is higher, followed by NN (0.933), BAG (0.930), GAM (0.923), DT (0.906), and LR (0.899), respectively. Fig. 5(b) shows the hydrographs representing the testing dataset's predicted and actual HHV values. It can be concluded from the figure that the prediction accuracy of VSGB is higher than that of the other designed ML. Table 3 depicts the statistical comparison among all the ML models.

The MSE value attained by the VSGB (Table 3) is 0.662, which is 44.8 %, 28.37 %, 21.4 %, 36.9 %, and 22.2 % lower than LR, GAM, BAG, DT, and NN. Furthermore, the AARD value for VSGB (2.827 %) is the lowest, followed by BAG (3.127 %), NN (3.356 %), DT (3.547 %), GAM (3.753 %), and LR (3.769%). In addition to this, LR attains the highest value of MAE (0.719) and RMSE (1.096), and VSGB has the lowest values of MAE (0.536) and RMSE (0.813) among all the designed ML techniques. The

above analysis shows the effectiveness of combining SR, GR, and BO. In VSGB, SR exhibits considerable robustness in high-dimensional domains and demonstrates a reduced susceptibility to overfitting, particularly with smaller datasets. The kernel method enables the model to capture complicated connections without explicitly expanding the feature space. Consequently, it is an excellent option for identifying non-linear patterns in data. GR offers probabilistic predictions and accounts for uncertainty in model outputs, enhancing accuracy in scenarios where data distributions are intricate or highly noisy. This effectively identifies trends in

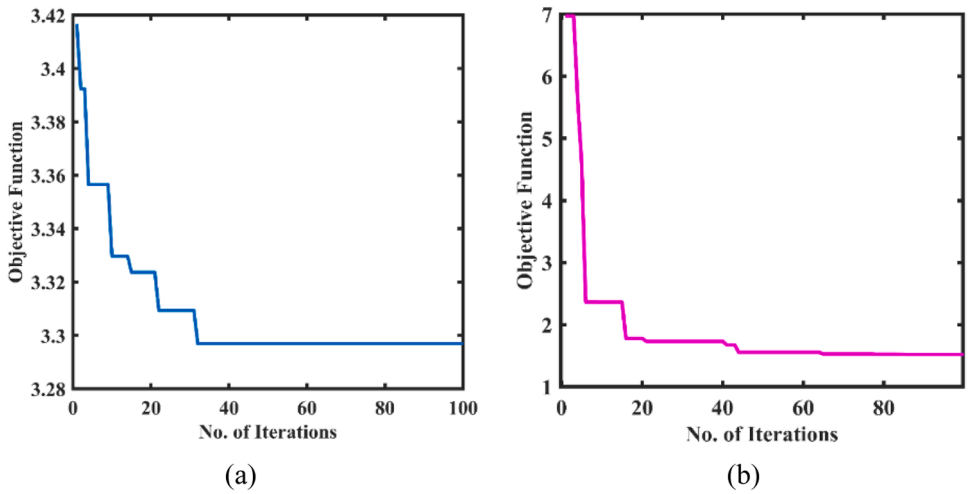
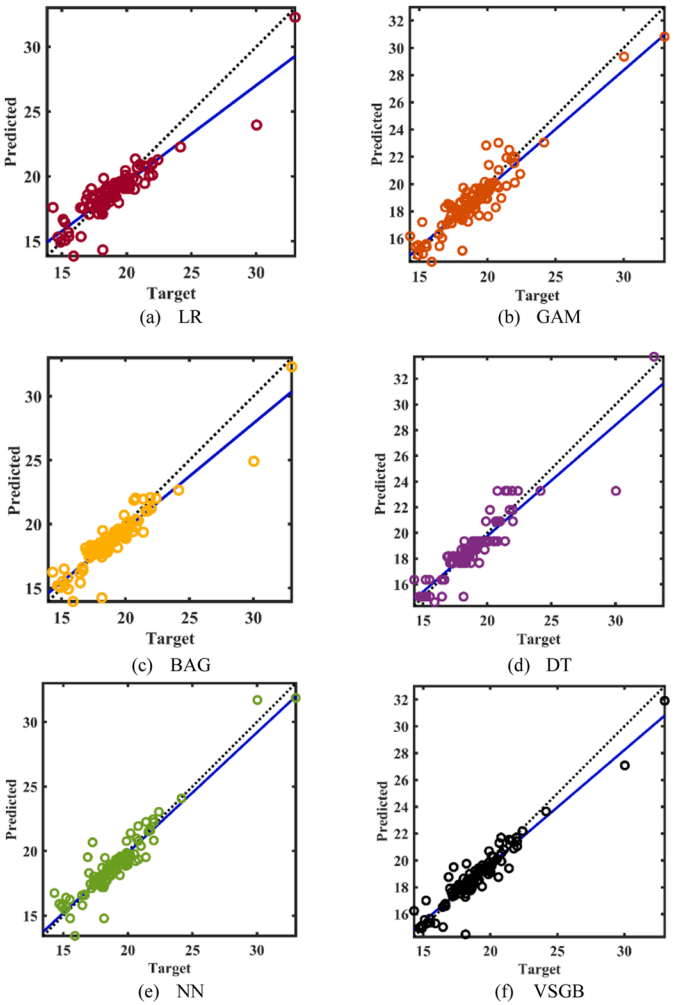
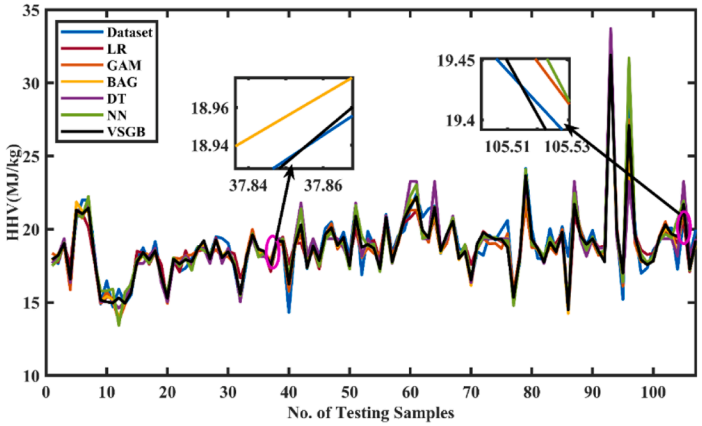


Fig. 4. Convergence plot for (a). VSGB (b). empirical equation for HHV prediction.



(a). Scatter plots between predicted and actual values for all the ML models



(b). Comparative Analysis of ML algorithms

Fig. 5. VSGB performance comparison on the testing dataset.

smaller datasets and addresses issues related to prediction uncertainty. It enhances the ensemble and aids the model in improving predictions by assessing the probability of each prediction's accuracy. This excludes SVR and BO, whose confidence predictions are not inherently regarded as a natural combination by default method that minimizes least square loss to rectify errors at each iteration and reduce biases in the model. Furthermore, BO sequentially trains the weak learner on prior errors, significantly diminishing predicting

Table 3
Comparative analysis of designed ML models based on performance indices.

Performance Indices	LR	GAM	BAG	DT	NN	VSGB
MSE	1.201	0.931	0.843	1.128	0.807	0.662
RMSE	1.096	0.965	0.918	1.062	0.898	0.813
MAE	0.719	0.713	0.602	0.691	0.620	0.536
AARD (%)	3.769	3.753	3.127	3.547	3.356	2.827

inaccuracy. BO is effective for ensemble issues that exhibit complex relationships, where residual errors may be overlooked by either SVR or GR in the overall fit.

3.1.1. Relevancy factor (r)

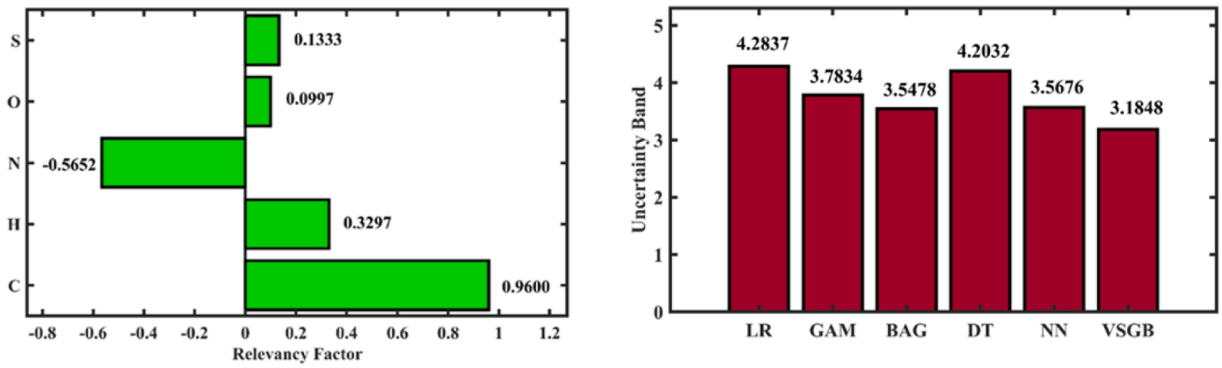
Using *r-analysis* is a dependable methodology for assessing the influence of the input attribute on the output. This quantifies the degree of influence that every input attribute exerts on the output. A higher *r-value* signifies a greater magnitude of influence that the input has on the output. The formula measures the relevancy factor (Hemmati-Sarapardeh et al., 2020).

$$r \quad (x_i, y) = \frac{\sum_{k=1}^{m} \left(x_{i,k} - x_{avg,i}\right)\left(y_k - y_{avg}\right)}{\sum_{k=1}^{m} \left(x_{i,k} - x_{avg,i}\right)^2 \quad \sum_{k=1}^{m} \left(y_k - y_{avg}\right)^2} \tag{20}$$

where

$x_{i,k}k^{th}$ values for the i^{th} input
$x_{avg,i}$Average value of i^{th} input
y_kPredicted output k^{th} value
y_{avg}Average output value

Fig. 6(a) shows the bar graph representation of the relevancy factor, calculated using Eq. (20). It is observed from the figure that the influence of C and N is higher on the HHV compared to H, O, and S. The *r* value estimated for C is 0.9600, followed by N (-0.5652), H (0.3297), S (0.1333) and O (0.0997), respectively. A precise prediction model of HHV derived from the final analysis is crucial for biochar producers, serving as a tool for informed decision-making, better production processes, cost savings, and improved environmental benefits. Real-time HHV projections enable operators to swiftly modify fuel compositions, moisture management, and other operational factors. This helps to optimize combustion efficiency, reduce fuel waste, and adhere to environmental emissions standards. It can also facilitate improved forecasting and planning to enable optimal energy production. Furthermore, uncertainty analysis was performed on all the designed ML techniques (Fig. 6(b)). The figure demonstrates that VSGB exhibits the lowest value of the uncertainty band, i.e., 3.1848, compared to other ML techniques for a 5 % significant level. Uncertainty analysis is essential for assessing the reliability of models forecasting the HHV generated by biochar from various biomass sources. The complete uncertainty interpretation of HHV value projections and its practicality in energy-generating projects justify its use. Significant feedstock variabilities in the energy generation project could discourage investors from returning to their conventional and predictable feedstocks, which are undoubtedly less sustainable and cost-effective for them. Energy generators may adopt a risk-averse approach, sustaining elevated biomass inventory levels, which escalates operational expenses and thus diminishes profitability. Since accurate and reliable HHV data is essential for regulatory compliance, considerable prediction discrepancies hinder compliance, leading to penalties. Project



(a). Relevancy factor

(b). Uncertainty band values for all the designed ML models

Fig. 6. (a). Relevancy factor. (b). Uncertainty band values for all the designed ML models.

developers should ensure that plans are sufficiently adaptable to accommodate actual HHV data in significant uncertainty. They must also guarantee that the systems employed can accommodate various biomasses or engage with several suppliers to mitigate feedstock fluctuation. Furthermore, the violin plots (Fig. 7) utilized in the investigation indicate that the VSGB model produces accurate forecasts of the HHV compared to the other models. Each violin plot illustrates the distribution of error range and density, reflecting the alignment of the model's predictions with the target values. The narrow error range of −2.2426 to –3.6597 for the VSGB model indicates limited variation and significant predictive accuracy. However, GAM ranges from −2.9336 to –3.0419, and BAG ranges from −1.9321 to –5.1184. In contrast, the other models have comparatively broader error distributions, suggesting more significant prediction variability and diminished precision in aligning with target HHV values.

The performance of VSGB is further compared with the ML techniques designed in the literature (Ighalo et al., 2020; Dai et al., 2021; Kotontung et al., 2022; Noushabdi et al., 2021; Yaka et al., 2022; Nieto et al., 2019; Samadi et al., 2021; Ighalo et al., 2022; Dashti et al., 2019), as shown in Fig. 8. The proposed VSGB has the lowest RMSE (0.813) and MSE (0.662) compared to the models developed in the literature. It can also be observed that the percentage of improvement achieved in the MSE is 6.7 %-84.5 % and in the RMSE is 3.4 %-90 %, respectively. The MAE attained by GARBF, ACO-ANFIS, and MNR are higher than the VSGB (0.536). In addition, the value of AARD% is lowest in the case of VSGB (2.827 %), which indicates the closeness of predicted HHV to the actual values. Moreover, it is crucial to recognize that the present comparison does not suggest the lack of effectiveness of the past research, as each of the preceding studies was carried out inside a particular environment, encompassing unique training and testing samples, among other variables. Policymakers and practitioners in the agricultural sector face numerous implications when applying machine learning models, such as VSGB, for biomass analysis and waste management. In conclusion, the enhanced accuracy is evidenced by reduced RMSE, MSE, MAE, and AARD% values in VSGB. It signifies potential reliability in forecasting the HHV of biomass, a critical aspect in assessing bioenergy potential. These findings indicate that VSGB may be pivotal for policymakers in formulating data-driven recommendations and standards for bioenergy production, facilitating enhanced utilization of agricultural by-products and waste

3.2. Empirical relationship estimation between input and output attributes using IWO

The preceding section proposes that VSGB predicts the HHV from the C, H, O, N, and S input attributes. The empirical relationship can also be used to estimate the HHV from inputs. In this work, an empirical equation is assumed which relates the C, H, O, N, and S to HHV as follows:

$$HHV = a1 * C + a2 * H + a3 * O + \left(\frac{a4 * N}{a5 * S}\right) \tag{21}$$

The IWO algorithm is used to calculate the values of the coefficients a1, a2, a3, a4, and a5. The convergence curve for the HHV empirical relationship with variations in the AARD% (on training samples) as the objective function is shown in Fig. 4(b). Table 4 compares the findings of the literature and the IWO-based empirical equation for HHV. Compared to others documented in the literature, the suggested equation yields the lowest value for the AARD%.

4. Conclusion

This article presents a design and implementation of VSGB for predicting HHV based on the elemental composition of C, H, O, N, and S obtained from the ultimate analysis. The VSGB is an aggregated sum of SR, GR, and BO-weighted values. In addition, the hyperparameter values of VSGB are evaluated using the IWO as follows: σ = 0.0001, Box Constraint = 67, Kernel Scale = 52, Epsilon = 0.1765, $\eta = 0.3276$, $w1 = 0.9984$, $w2 = 0.9388$, and $w3 = 0.4301$. The performance of the VSGB is evaluated and compared with several other models, including LR, GAM, DT, BAG, and NN, using several performance measures. The findings indicate that the VSGB technique achieves the lowest MSE value of 0.662 and AARD% of 2.827 when compared to other ML techniques that were implemented. Moreover, an equation that predicts the HHV based on the elemental composition of C, H, O, N, and S is formulated using the IWO algorithm. The proposed relationship yielded a lower value of AARD% (3.59%) than other equations documented in the existing literature. In addition to this, research questions have been answered about the above work as follows:

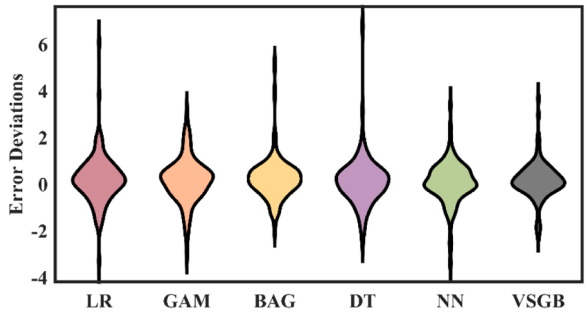
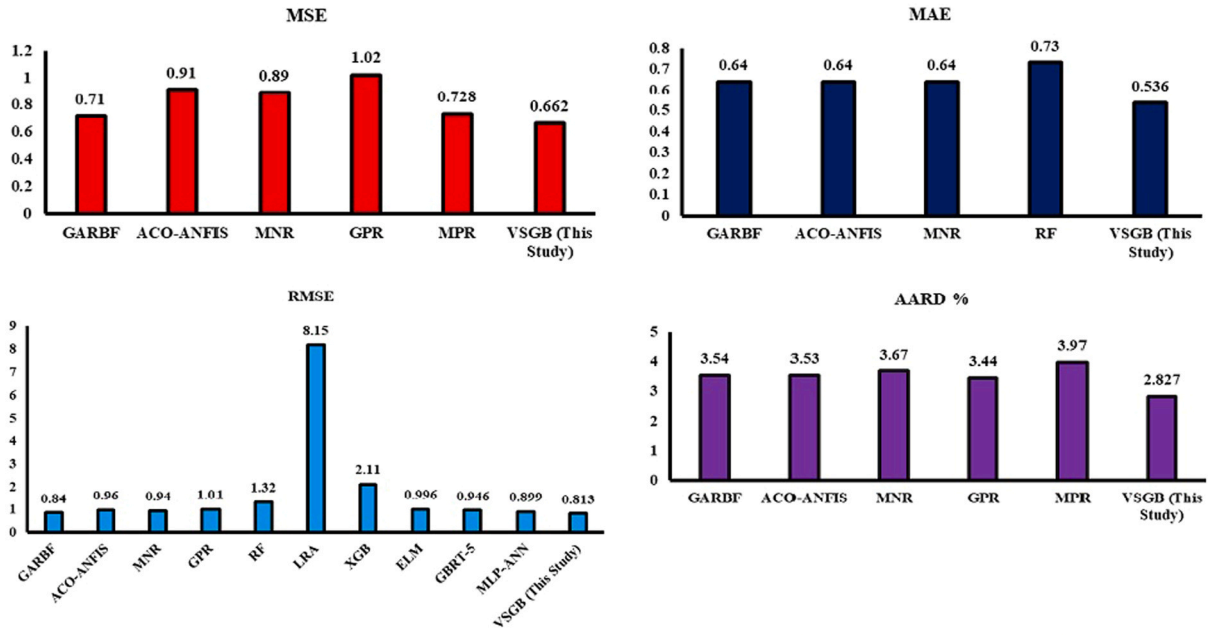


Fig. 7. Error distribution using violin plot.



GARBF, ACO-ANFIS, MNR (Noushabadi *et.al* 2021); GPR (Yaka *et.al* 2022); RF(Nieto *et.al* 2022): LRA (Ighalo *et.al* 2020); XGB (Katongtung *et.al* 2022); ELM (Dai *et.al* 2021); GBRT-5 (Samadi *et.al* 2019); MLP-ANN (Ighalo *et.al* 2022); MPR (Dashti *et.al* 2019)

Fig. 8. Comparative Analysis w.r.t different performance indices of VSGB with ML models given in the literature.

Table 4
IWO-HHV predictive equation comparison with the literature.

Predictive Equations	AARD %	Ref
$HHV = -0.0769C - 0.3110O + 35.8357$	13.94	(Boumanchar et al., 2019)
$HHV = 0.6342H - 0.2142O + 0.8943S + 23.7283$	11.33	(Boumanchar et al., 2019)
$HHV = C + H + S + \frac{174}{C-63.3} + \frac{0.399}{O-52} + \frac{0.0727}{C+H+53.4S-HS-1.1N} - 23.7$	27.79	(Boumanchar et al., 2019)
$HHV = -3.440 + 0.517(C+N) - 0.433(H+N)$	6.16	(Xing et al., 2019)
$HHV = -5.290 + 0.493C + 5.052(H)^{-1}$	7.58	(Xing et al., 2019)
$HHV = 5.736 + 0.0006C^2$	12.68	(Xing et al., 2019)
$HHV = (1.59C^2 + 154.5C + 7464) * 10^{-3}$	4.76	(Callejon-Ferre et al., 2011)
$HHV = 3.55C^2 - 232C - 2230H + 51.2C * H + 131N + 20600$	4.28	(García et al., 2014)
$HHV = 0.367C + \frac{53.883O}{(2.31C^2 - 3.299)} + \frac{(C * H - 115.971)}{\left(10.472 * H + 0.129C * O\left(\frac{-91.531}{(35.299+N)}\right)\right)} + \frac{232.698}{(77.545+S)}$	3.65	(Friedl et al., 2005)
$HHV = -0.3516C + 1.1625H - 0.1109O + 0.0628N + 0.10465S$	5.06	(Ghugare et al., 2014)
$HHV = 357.8C + 1135.6H + 54.9N - 85.4O + 119.5S - 974$	4.83	(Mason and Gandhi, 1980)
$HHV = 144.4C + 610.2H - 65.9O + 0.39C^2$	5.53	(Cordero et al., 2001)
$HHV = 78.31C + 359.32\left(H - \frac{O}{8}\right) + 22.12S + 11.87O + 5.78N$	5.95	(Kathiravale et al., 2003)
$HHV = -0.8738 * N * H^{-1.3101} - 0.1583 * C * O^{0.3497} + 0.3856 * C(H * O)^{0.1462} + 2.1436\left(\frac{H}{O}\right)^{-0.3846} + 0.1076 * C * H^{-0.3846} + 0.1098 * N * S - 11.2794\left(\frac{H}{C}\right)$	3.61	(Noushabadi et al., 2021)
$HHV = 0.3716 * C + 0.0586 * H + 0.0207 * O + \left(\frac{0.1850 * N}{0.9857 * S}\right)$	**3.59**	**This Study**

What impact would accurate HHV prediction have on energy regulations and advancements in bioenergy?

Answer - The precise prediction of HHV, as achieved by the ML models, would facilitate the identification of the most efficient biomass sources for energy. Policymakers can utilize these projections to actively advocate for high-efficiency biomass feedstocks. Furthermore, it can optimize energy production to minimize waste in energy generation. This would help to make rules that encourage the long-term growth and processing of biomass to lower our reliance on fossil fuels.

In what ways do advanced machine learning models enhance the predictive accuracy of HHV in various biomass materials for sustainable energy applications?

Answer- Advanced ML algorithms can enhance biomass HHV predictions by analyzing data patterns. These models calculate HHV based on moisture, ash, and carbon content. They are adaptable and boost biomass reliability. Precise HHV prediction aids in assessing fuel quality, reducing waste, and optimizing energy production.

Future scope- The proximate analysis will be supplemented by the ultimate analysis for HHV predictions. Empirical validation using many biomass datasets from different categories will assess the model's generalisability. A comparative analysis with new machine learning approaches will be conducted to evaluate the performance of the proposed model. The model will be enhanced for extensive bioenergy forecasting by integrating deep learning techniques. Deep learning models will identify intricate correlations and non-linear patterns in big datasets, enhancing predictive accuracy. Moreover, future studies may examine the economic and environmental consequences of employing ML in bioenergy production. A detailed analysis of the relevancy factor will also be investigated for different biomass types based on their elemental composition.

Limitation- The complexity of the amalgamation of SR, GR, and BO values may restrict adaptability. The sensitivity of hyperparameters indicates that performance is dependent upon their configuration. Advanced approaches like IWO are computationally expensive; hence, executing real-time or large-scale implementations may be challenging. Variations in biomass outside the model's parameters and environmental conditions may affect the model's accuracy. The study focused on HHV prediction without considering other essential properties like ash content or lignin.

CRediT authorship contribution statement

Nikhil Pachauri: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tae Jong Choi:** Writing – original draft, Resources, Funding acquisition. **Chang Wook Ahn:** Writing – review & editing, Supervision, Project administration.

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.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00214326). The correspondence should be addressed to both Dr. Chang Wook Ahn and Dr. Tae Jong Choi since they contributed equally to this work.

Authorship confirmation

All authors have read and approved the final version of the manuscript and agree to its submission to Environmental Technology & Innovation.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.eti.2024.104012.

Data availability

Data will be made available on request. The data that support the findings of this study are available on request

References

59⟨https://biochar-international.org/fy24-impact-snapshot/⟩. ⟨https://biochar-international.org/fy24-impact-snapshot/⟩.
60⟨https://www.irena.org/media/Files/IRENA/Agency/Publication/2019/Jun/IRENA_G20_climate_sustainability_2019.pdf⟩ ⟨https://www.irena.org/media/Files/IRENA/Agency/Publication/2019/Jun/IRENA_G20_climate_sustainability_2019.pdf⟩.

Adeniyi, A.G., Ighalo, J.O., 2020. Computer-aided modeling of thermochemical conversion processes for environmental waste management. In: Hussain, C.M. (Ed.), Handbook of Environmental Materials Management. Springer Nature, Cham, pp. 1–16.

Adeniyi, A.G., Ighalo, J.O., Abdulsalam, A., 2019. Modelling of integrated processes for the recovery of the energetic content of sugarcane bagasse. Biofuels Bioprod. Biorefin (4), 1057–1067.

Ahmaruzzaman, M., 2008. Proximate analyses and predicting HHV of chars obtained from cocracking of petroleum vacuum residue with coal, plastics, and biomass. Bioresour. Technol. 99 (11), 5043–5050.

Barisal, A.K., Prusty, R.C., 2015. Large-scale economic dispatch of power systems using oppositional invasive weed optimization. Appl. Soft Comput. 29, 122–137.

Boumanchar, I., Charafeddine, K., Chhiti, Y., Alaoui, F.E.M.H., Sahibed-dine, A., Bentiss, F., Jama, C., Bensitel, M., 2019. Biomass higher heating value prediction from ultimate analysis using multiple regression and genetic programming. Biomass-.-. Convers. Biorefin. 1–11.

Brandić, Ivan, Pezo, Lato, Bilandžija, Nikola, Peter, Anamarija, Šurić, Jona, Voća, Neven, 2022. Artificial neural network as a tool for estimation of the higher heating value of miscanthus based on ultimate analysis. Mathematics 10 (20), 3732.

Brandić, Ivan, Voća, Neven, Gunjača, Jerko, Lončar, Biljana, Bilandžija, Nikola, Peter, Anamarija, Šurić, Jona, Pezo, Lato, 2024. "Biomass higher heating value prediction machine learning insights into ultimate, proximate, and structural analysis datasets. Energy Sources Part A Recovery Util. Environ. Eff. 46 (1), 2842–2854.

Callejon-Ferre, Velazquez-Martí, B., Lopez-Martínez, J., Manzano-Agugliaro, F., 2011. Greenhouse crop residues: energy potential and models for the prediction of their higher heating value. Renew. Sustain. Energy Rev. 15 (2), 948–955.

Cao, Hongliang, Xin, Ya, Yuan, Qiaoxia, 2016. Prediction of biochar yield from cattle manure pyrolysis via least squares support vector machine intelligent approach. Bioresour. Technol. 202, 158–164.

Channiwala, S., Parikh, P., 2002. A unified correlation for estimating HHV of solid, liquid and gaseous fuels. Fuel 81 (8), 1051–1063.

Chen, Guanxing, Jiang, Xuefei, Lv, Qiujie, Tan, Xiaojun, Yang, Zihuan, Chen, Calvin Yu-Chian, 2022. "VAERHNN: voting-averaged ensemble regression and hybrid neural network to investigate potent leads against colorectal cancer. Knowl. Based Syst. 257, 109925.

Chen, D., Zhao, J., Qin, S., 2023. SVM strategy and analysis of a three-phase quasi-Z-source inverter with high voltage transmission ratio. Sci. China Technol. Sci. 66 (10), 2996–3010.

Cordero, T., Marquez, F., Rodriguez-Mirasol, J., Rodriguez, J., 2001. Predicting heating values of lignocellulosics and carbonaceous materials from proximate analysis. Fuel 80 (11), 1567–1571.

Dai, Zuocai, Chen, Zhengxian, Selmi, Abdellatif, Jermsittiparsert, Kittisak, Denić, Nebojša.M., Nešić, Zoran, 2021. Machine learning prediction of higher heating value of biomass. Biomass-.-. Convers. Biorefinery 1–9.

Dashti, Amir, Noushabadi, Abolfazl Sajadi, Raji, Mojtaba, Razmi, Amir, Ceylan, Selim, Mohammadi, Amir H., 2019. Estimation of biomass higher heating value (HHV) based on the proximate analysis: smart modeling and correlation. Fuel 257, 115931.

Demiral, İ., 2009. Atilgan NG and Şensöz S, Production of biofuel from soft shell of pistachio (*Pistacia vera* L.). Chem. Eng. Commun. 196 (1–2), 104–115.

Dubey, Richa, Guruviah, Velmathi, 2023. Predictive modeling of higher heating value of biomass using ensemble machine learning approach. Arab. J. Sci. Eng. 48 (7), 9329–9338.

Fei, R., Guo, Y., Li, J., Hu, B., Yang, L., 2023. An improved BPNN method based on probability density for indoor location. IEICE Trans. Inf. Syst. 106 (5), 773–785.

Friedl, E., Padouvas, H., Rotter, K., Varmuza, 2005. Prediction of heating values of biomass fuel from elemental composition. Anal. Chim. Acta 544 (1-2), 191–198.

García, R., Pizarro, C., Lavín, A.G., Bueno, J.L., 2014. Spanish biofuels heating value estimation. Part I: ultimate analysis data. Fuel 117, 1130–1138.

Ghugare, S., Tiwary, S., Elangovan, V., Tambe, S., 2014. Prediction of higher heating value of solid biomass fuels using artificial intelligence formalisms. Bio Energy Res. 7 (2), 681–692.

Güleç, Fatih, Pekaslan, Direnc, Williams, Orla, Lester, Edward, 2022. Predictability of higher heating value of biomass feedstocks via proximate and ultimate analyses–a comprehensive study of artificial neural network applications. Fuel 320, 123944.

Guo, J., Liu, Y., Zou, Q., Ye, L., Zhu, S., Zhang, H., 2023. Study on optimization and combination strategy of multiple daily runoff prediction models coupled with physical mechanism and LSTM. J. Hydrol. 624, 129969.

Hemmati-Sarapardeh, Abdolhossein, Larestani, A., Menad, N.A., Hajirezaie, S., 2020. Applications of artificial intelligence techniques in the petroleum industry. Gulf Professional Publishing.

Ighalo, Joshua O., Adeniyi, Adewale George, Marques, Gonçalo, 2024. Application of artificial neural networks in predicting biomass higher heating value: an early appraisal. Energy Sources Part A Recovery Util. Environ. Eff. 46 (1), 15117–15124.

Ighalo, Joshua O., Igwegbe, Chinenye Adaobi, Adeniyi, Adewale George, 2022. Multi-layer perceptron artificial neural network (MLP-ANN) prediction of biomass higher heating value (hhv) using combined biomass proximate and ultimate analysis data. Model. Earth Syst. Environ. 8 (3), 3177–3191.

Ighalo, Joshua, O., Adeniyi, Adewale George, Marques, Goncalo, 2020. Application of linear regression algorithm and stochastic gradient descent in a machine-learning environment for predicting biomass higher heating value. Biofuels Bioprod. Bioref. 14 (6), 1286–1295.

Jakšić, Olga, Jakšić, Zoran, Guha, Koushik, Silva, Ana G., Laskar, Naushad Manzoor, 2023. Comparing artificial neural network algorithms for prediction of higher heating value for different types of biomass. Soft Comput. 27 (9), 5933–5950.

Kandpal, Saurav, Tagade, Ankita, Sawarkar, Ashish N., 2024. "Critical insights into ensemble learning with decision trees for the prediction of biochar yield and higher heating value from pyrolysis of biomass. Bioresour. Technol. 411, 131321.

Kartal, Furkan, Özveren, U.ğur, 2022. Investigation of the chemical exergy of torrefied biomass from raw biomass by means of machine learning. Biomass-.-. Bioenergy 159, 106383.

Kathiravale, S., Yunus, M.N.M., Sopian, K., Samsuddin, A.H., 2003. R. Rahman, Modeling the heating value of municipal solid Waste*. Fuel 82 (9), 1119–1125.

Katongtung, Tossapon, Onsree, Thossaporn, Tippayawong, Nakorn, 2022. Machine learning prediction of biocrude yields and higher heating values from hydrothermal liquefaction of wet biomass and wastes. Bioresour. Technol. 344, 126278.

Kaya, Esma Yeliz, Ali, Imtiaz, Ceylan, Zeynep, Ceylan, Selim, 2024. "Prediction of higher heating value of hydrochars using Bayesian optimization tuned Gaussian process regression based on biomass characteristics and process conditions. Biomass-.-. Bioenergy 180, 106993.

Kocer, Abdulkadir, 2024. Prediction of the higher heating values of biomass using machine learning methods based on proximate and ultimate analysis. J. Mech. Sci. Technol. 38 (3), 1569–1574.

Liu, Xiangjie, Zhang, Xin, Charoenkal, Khantaphong, Yuan, Qiaoxia, Cao, Hongliang, 2024. Hydrothermal bio-oil yield and higher heating value of high moisture and lipid biomass: machine Learning modeling and feature response behavior analysis. J. Energy Inst., 101859

Mason, D., Gandhi, K., 1980. Formulas for Calculating the Heating Value of Coal and Coal Char: Development, Tests, and Uses. Institute of Gas Technology, Chicago, IL (USA).

Mehrabian, Ali Reza, Lucas, Caro, 2006. A novel numerical optimization algorithm inspired from weed colonization. Ecol. Inform. 1 (4), 355–366.

Mu, Lin, Wang, Zhen, Wu, Di, Zhao, Liang, Yin, Hongchao, 2022. Prediction and evaluation of fuel properties of hydrochar from waste solid biomass: machine learning algorithm based on proposed PSO–NN model. Fuel 318, 123644.

Nieto, García, P.J., García–Gonzalo, E., Sánchez Lasheras, F., Paredes–Sánchez, J.P., Riesgo Fernández, P., 2019. Forecast of the higher heating value in biomass torrefaction by means of machine learning techniques. J. Comput. Appl. Math. 357, 284–301.

Noushabadi, Abolfazl Sajadi, Dashti, Amir, Ahmadijokani, Farhad, Hu, Jinguang, Mohammadi, Amir H., 2021. Estimation of higher heating values (HHVs) of biomass fuels based on ultimate analysis using machine learning techniques and improved equation. Renew. Energy 179, 550–562.
Onsree, Thossaporn, Tippayawong, Nakorn, Phithakkitnukoon, Santi, Lauterbach, Jochen, 2022. Interpretable machine-learning model with a collaborative game approach to predict yields and higher heating value of torrefied biomass. Energy 249, 123676.
Pachauri, Nikhil, Ahn, Chang Wook, 2023. Weighted aggregated ensemble model for energy demand management of buildings. Energy 263, 125853.
Pachauri, Nikhil, Ahn, Chang Wook, 2022. Regression tree ensemble learning-based prediction of the heating and cooling loads of residential buildings. In: Building Simulation, 15. Tsinghua University Press, Beijing, pp. 2003–2017.
Parikh, Jigisha, Channiwala, S.A., Ghosal, G.K., 2005. A correlation for calculating HHV from proximate analysis of solid fuels. Fuel 84 (5), 487–494.
Puig-Arnavat, M., Bruno, J.C., Coronas, A., 2010. Review and analysis of biomass gasification models. Renew. Sustain Energy Rev. 14, 2841e51.
Samadi, Seyed Hashem, Ghobadian, Barat, Nosrati, Mohsen, 2021. Prediction of higher heating value of biomass materials based on proximate analysis using gradient boosted regression trees method. Energy Sources Part A Recovery Util. Environ. Eff. 43 (6), 672–681.
Schulz, Eric, Speekenbrink, Maarten, Krause, Andreas, 2018. A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions. J. Math. Psychol. 85, 1–16.
Shi, Honghong, Mahinpey, Nader, Aqsha, Aqsha, Silbermann, Rico, 2016. Characterization, thermochemical conversion studies, and heating value modeling of municipal solid waste. Waste Manag. 48, 34–47.
Taki, Morteza, Rohani, Abbas, 2022. Machine learning models for prediction the higher heating value (HHV) of municipal solid waste (MSW) for waste-to-energy evaluation. Case Stud. Therm. Eng. 31, 101823.
Timilsina, Sharma, M., Sen, S., Uprety, B., Patel, V.B., Sharma, Prateek, Sheth, Pratik N., 2024. Prediction of HHV of fuel by machine learning algorithm: interpretability analysis using shapley additive explanations (SHAP). Fuel 357, 129573.
Vargas-Moreno, J., Callejón-Ferre, A., Pérez-Alonso, J., Velázquez-Martí, B., 2012. A review of the mathematical models for predicting the heating value of biomass materials. Renew. Sustain Energy Rev. 16 (5), 3065–3083.
Wang, Minghong, Xie, Yingpu, Gao, Yong, Huang, Xiaohong, Chen, Wei, 2024. Machine learning prediction of higher heating value of biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. 395, 130364.
Wen, X., Luo, K., Jin, H.H., Fan, J.R., 2017. Large eddy simulation of piloted pulverized coal combustion using extended flamelet/progress variable model. Combust. Theor. Model 21, 925e53.
Xing, Jiangkuan, Luo, Kun, Wang, Haiou, Gao, Zhengwei, Fan, Jianren, 2019. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. Energy 188, 116077.
Yaka, Havva, Insel, Mert Akin, Yucel, Ozgun, Sadikoglu, Hasan, 2022. A comparison of machine learning algorithms for estimation of higher heating values of biomass and fossil fuels from ultimate analysis. Fuel 320, 123971.
Zhu, C., Wang, M., Guo, M., Deng, J., Du, Q., Wei, W., Mohebbi, A., 2024. An innovative process design and multi-criteria study/optimization of a biomass digestion-supercritical carbon dioxide scenario toward boosting a geothermal-driven cogeneration system for power and heat. Energy 292, 130408.
Zhu, Xingwei, Yang, Guozhu, 2022. Study on HHV prediction of municipal solid wastes: a machine learning approach. Int. J. Energy Res. 46 (3), 3663–3673.