Digital Chemical Engineering 12 (2024) 100159

Contents lists available at ScienceDirect

Digital Chemical Engineering

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

Original Article

Comparative studies of machine learning models for predicting higher heating values of biomass

Adekunle A. Adeleke [a], Adeyinka Adedigba [b], Steve A. Adeshina [c], Peter P. Ikubanni [d], Mohammed S. Lawal [e], Adebayo I. Olosho [f], Halima S. Yakubu [g], Temitayo S. Ogedengbe [a], Petrus Nzerem [h], Jude A. Okolie [i,*]

[a] Department of Mechanical Engineering, Nile University of Nigeria, Abuja, Nigeria
[b] Department of Mechatronics Engineering, Federal University of Technology, Minna, Nigeria
[c] Department of Computer Engineering, Nile University of Nigeria, Abuja, Nigeria
[d] Department of Mechanical Engineering, Landmark University, Omu Aran, Nigeria
[e] Department of Mechanical Engineering, Air Force Institute of Technology, Kaduna, Nigeria
[f] Department of Industrial Chemistry, University of Ilorin, Ilorin, Nigeria
[g] Department of Electrical Engineering, Nile University of Nigeria, Abuja, Nigeria
[h] Department of Petroleum and Gas Engineering, Nile University of Nigeria, Abuja, Nigeria
[i] Gallogly College of Engineering, University of Oklahoma, Norman United States

ARTICLE INFO

Keywords:
Biomass materials
Energy crops
Machine learning
Ultimate analysis
Proximate analysis
Higher heating values

ABSTRACT

This study addresses the challenge of efficiently determining the higher heating value (HHV) of biomass, a crucial parameter in large-scale biomass-based energy systems. The conventional method of measuring HHV using an oxygen bomb calorimeter is time-consuming, expensive, and less accessible to researchers, particularly in developing nations. To overcome these limitations, we employed four machine learning (ML) models, namely Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). These models were developed by using proximate and ultimate analysis parameters as input features. Up to 200 datasets were compiled from literature and used for the ML models. Our results demonstrate the effectiveness of all ML models in accurately predicting the HHV of biomass materials. Notably, the XGBoost model exhibited superior performance with the highest R-squared (R^2) values for both training (0.9683) and test datasets (0.7309), along with the lowest root mean squared error (RSME) of 0.3558. Key influential input features identified for HHV prediction include carbon (C), volatile matter (Vm), ash, and hydrogen (H). Consequently, this research provides a reliable alternative for predicting HHV without the need for costly and time-intensive experimental measurements, facilitating broader accessibility in biomass energy research.

1. Introduction

The rate at which fossil fuel depletes and the environmental challenges associated with its usage have become the center of attraction among various researchers across the globe (Adeleke et al., 2019). The issue of greenhouse gas emissions and global warming is a key challenge associated with the continuous use of fossil fuels as the main source of energy (Adeleke et al., 2021). To avoid a complete breakdown of the ozone layer and disruption of the climatic conditions, alternative fuels and sources of energy have become imperative. Renewable sources such as hydro, solar, wind, geothermal, and biomass have been attractive and there is an increasing demand for them (Taki and Rohani, 2022). The use of biomass as a source of energy and for the production of sustainable industrial materials is now an area of interest to researchers. This is because of its availability, renewability, and carbon neutrality (Odusote et al., 2019; Gulec et al., 2022). The rapidly growing population, especially in developing countries, as well as the increasing energy demand necessitates an additional clean, sustainable, and renewable energy source. Biomass is one of these renewable sources that has similar properties to coal though it has higher volatile matter, low energy contents and it is hydrophobic. However, the carbon neutrality, low ash, and high carbon content make it a preferred source of energy to coal.

* Corresponding author.
E-mail addresses: adekunle.adeleke@nileuniversity.edu.ng (A.A. Adeleke), jude.a.okolie@gmail.com (J.A. Okolie).

https://doi.org/10.1016/j.dche.2024.100159
Received 4 January 2024; Received in revised form 21 May 2024; Accepted 26 May 2024
Available online 29 June 2024

Biomass can partially or totally replace coal in energy generation and metallurgical applications (Doddapaneni and Kikas, 2023). It can also be effective for the production of bio-oil that can replace petroleum-based diesel (Verger et al., 2022). To obtain energy from biomass, thermochemical conversion methods such as gasification and pyrolysis, or biochemical conversion methods such as compositing can be used. The design and construction of these energy systems depend on a major characteristic of biomass which is the calorific value or higher heating value (Ayeleru et al., 2020). Thus, a need to have a deep understanding and knowledge of the heating value of biomass for effective usage during combustion and operation of such an energy system.

Higher heating value (HHV) is very crucial to the thermal combustion of biomass. HHV is referred to as the heat released when a unit mass of biomass is completely combusted. Adeleke et al. (2021) opined that the carbon, hydrogen, oxygen, fixed carbon, and ash content of any fuel are indicators of its HHV. Significantly, carbon, hydrogen, and oxygen are crucial variables influencing the combustion process for heat production (Vallejo et al., 2020). The accurate determination of biomass HHV value is important for the designing, constructing, and effectively managing its waste-to-energy system (Scarlat et al., 2019). HHV of biomass is experimentally measured using a bomb calorimeter by following some stipulated ASTM, IS, and BS standards. The challenge with developing nations is that the equipment is rarely available due to its cost. For instance, few bomb calorimeters can be found in Nigeria and the majority are commercially used which comes at a huge price for one sample. The use of fast and cost-effective methods to estimate the HHV of biomass for waste-to-energy systems is now attractive to researchers with limited or no access to bomb calorimeters. This approach integrates the use of machine learning (ML) techniques to assess the multi-dimensional relationships of biomass properties (proximate, ultimate, and HHV).

ML is a sub-branch of artificial intelligence (AI) that can automatically learn and recognize patterns from a dataset without being explicitly programmed (Onsree and Tippayawong, 2021; Adeshina and Adedigba, 2022). Data-driven ML approach has become an integral part of engineering and scientific research across the globe. Several researchers have adopted ML in energy-related studies. Some used a single model (Al-Janabi et al., 2020; Strubell et al., 2020) while few adopted multiple models in their studies (Yan et al., 2023; Adeleke et al., 2023). Onsree and Tippayawong (2021) applied ML to predict yields of solid products from biomass torrefaction. Several ML algorithms were evaluated with ten-fold cross-validation. It was concluded that the ML provided a better perspective to understand the multidimensional relationships in the biomass torrefaction process. Gulec et al. (2022) carried out a comprehensive study of artificial neural network (ANN) applications to predict HHV of biomass feedstock via proximate and ultimate analyses. The work presented how thirteen different algorithms, hidden layers, dataset, and randomization of datasets affect the HHV prediction of biomass. The ANN models trained by combining proximate and ultimate analyses were more accurate than those trained using proximate or ultimate analysis only. Proximate analyses yield moisture content, ash content, fixed carbon, and volatile matter content while the ultimate contents include carbon, hydrogen, oxygen, nitrogen, and sulfur (Adeleke et al., 2019; 2020). Both are to be combined to have an effective prediction of the HHV of biomass. Ighalo et al. (2022) also used a multilayer perceptron artificial neural network to predict the HHV of biomass by combining proximate and ultimate analyses. Two hundred and one (201) data points were utilized for both proximate and ultimate analyses to predict HHV. High R^2 values were obtained for both testing and validation sets which showed the good estimation and generalization capacity of the ANN model. Yan et al. (2023) predicted the HHV of sewage sludge via an ANN based on proximate and ultimate analyses. Back-propagation neural network models based on proximate and ultimate analyses were developed to predict HHV of the sewage sludge. It was reported that the model had good accuracy with a regression coefficient of 0.979 and 0.975 for the training and test groups, respectively.

From the foregoing discussion, the use of ANN for predicting HHV dominates the research as demonstrated by the works of Güleç et al. (2022).However, ANN poses some challenges: (i) model complexity which makes it easily overfit smaller dataset, such as common in this research, and (ii) lack of model explainability i.e. it is difficult to know which feature are important in arriving at the accurate prediction. While there have been several studies that used data-driven ML methods for the prediction of HHV values, most of the studies either use proximate analysis or ultimate analysis data. For instance, Dai et al. (2021) developed ML model for HHV prediction based on only ultimate analysis. Similarly, Xing et al. (2019) developed a robust ML model with ultimate analysis data. Another issue with HHV predictions is the lack of robust dataset that comprises of several categories of biomass materials. Afolabi et al. (2022) study was one of the first to develop a ML model for HHV prediction that integrates different classes of biomass as input variables. Regardless, the use of proximate and ultimate analysis data as well as incorporating a robust dataset comprising of agricultural residues, industrial waste, woody biomass, and energy crops is an area that requires further attention. To address the knowledge gaps, we explore a comprehensive ML model in this work. We consider the support vector machines (SVM) due to its ability to achieve higher prediction accuracy with minimum model complexity, thanks to its kernel trick. In addition, we presented three tree-based models due to their explainability which is useful for gaining additional insight into the features used in predicting HHV. We utilized proximate and ultimate analysis data as well as a robust dataset that captures several categories of lignocellulosic biomass. We also compare the performance of our model with empirical models and other data-driven ML models to further verify its accuracy.

2. Methodology

2.1. Overview of the dataset

The dataset which was used contains information about proximate and ultimate analysis as well as the HHV of two hundred (200) biomasses. These data were gathered from existing literature as listed in the supplementary material. The data was obtained from various biomass which includes forest residue, energy crops, sawmill wastes, and agricultural lignocellulosic wastes. It should be mentioned that one of the motivations of this study is to capture several ranges of biomass materials representing different classes of biomass. We do not want one class to be overrepresented, this explains why we kept the data low at 200.

Nine variables that capture both the proximate and ultimate analysis of biomass were considered and are presented in Table 1. Furthermore, the data were split into proximate content (volatile matter (VM), ash content (ash), and fixed carbon (FC)) and ultimate contents (carbon, hydrogen, sulphur, nitrogen, and oxygen) while the target was biomass HHV value. The uncertainty of the data for each variable obtained from the literature was within 0.05. Details of the procedure utilized in this study are presented in Fig. 1.

Table 1
Inputs and output variables.

S/N	Features	ID	Range
Ultimate contents (wt.%) air-dried basis			
1	Carbon content	C	26.70 - 62.10
2	Hydrogen content	H	0.27 - 8.06
3	Nitrogen	N	0.02 - 4.33
4	Sulphur	S	0 - 1.00
5	Oxygen	O	28.80 – 70.05
Proximate contents (wt.%) air-dried basis			
6	Volatile matter content	VM	59.20 – 91.40
7	Ash content	Ash	0.10 – 67.75
8	Fixed carbon content	FC	2.10 – 35.80
Target			
9	Higher heating value (MJ/kg)	HHV	13.24 – 25.79

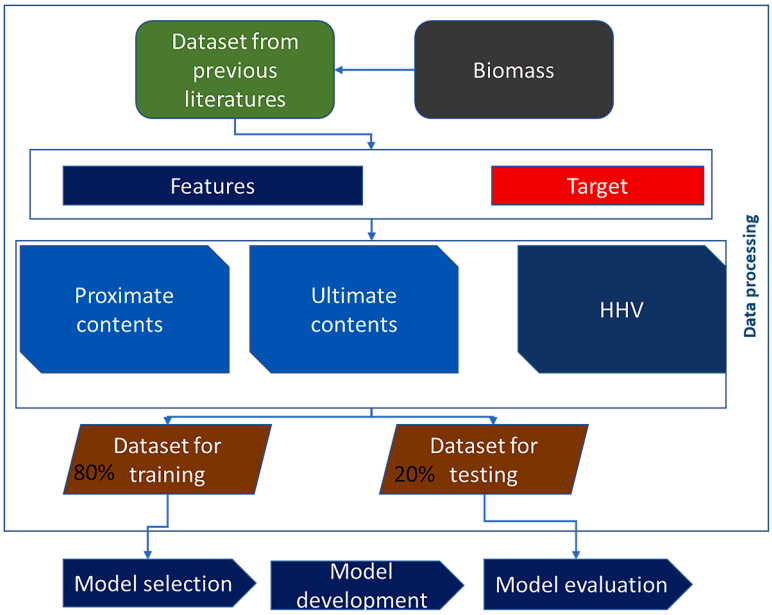

Fig. 1. Framework of the methodology.

2.2. Description of the machine learning models

In this work, four ML models known for their good performance on prediction tasks were used. A support vector machine (SVM) method and three tree-based algorithms – decision tree, random forest, and extended gradient boost – were selected for the task. These models are selected based on their promising performance in prior studies related to thermochemical conversion processes (Umenweke et al., 2022). A brief description of each model is presented in this section. However, interested readers are referred to excellent review by Umenweke et al. (2022) for detailed information about each ML model.

2.2.1. Support vector regression

This is a kind of support vector machine (SVM) designed explicitly for regression problems, as against the classification problem native to SVM. Given training data of the form $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \in \chi = \Re^d$, where χ denotes the distribution of the input space. The goal of a support vector regressor (SVR) is to find a function $f(x)$ that takes input x_i and produces an output $\widehat{y}_i$ that has at most ε deviation from the target output y_i and at the same time, the function $f(x)$ must be as flat as possible. The flatness of the function imposes a constraint on the parameters of $f(x)$ from becoming too large, that is, $f(x)$ is flat when the norm of w, $\left(\| w\text{AptCommand2016;}^2 \right)$ is as small as possible.

Finding $f(x)$ that perfectly fits the data with only ε allowable deviation, and that minimizes the norm of w is a convex optimization problem that is not easy to solve, but when some additional margin of error ξ is allowed, we can develop a better approximator. The SVR used for many practical applications employs the Lagrange multipliers to solve the dual optimization problem of minimizing the deviation from $(\varepsilon + \xi)$ and $\left(\| w\text{AptCommand2016;}^2 \right)$ as can be seen in Eq.(1).

$$L: = \frac{1}{2}\| w \|^2 + \lambda \sum_{n}^{i=1} \left(\xi_i + \xi_i^*\right) - \sum_{n}^{i=1} \left(\eta_i \xi_i + \eta_i^* \xi_i^*\right) - \sum_{n}^{i=1} \alpha_i (\varepsilon + \xi_i - y_i + f(x_i)) - \sum_{n}^{i=1} \alpha_i^* \left(\varepsilon + \xi_i^* + y_i - f(x_i)\right) \quad (1)$$

Where L is the Lagrangian and $\eta_i,\ \eta_i^*,\ \alpha_i,\ \alpha_i^*$ are the Lagrange multipliers. Hence, the optimization problem can be written as Eq. (2)-(3) i.e.:

Maximize

$$\begin{cases} -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)\langle x_i, x_j \rangle \\ -\varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i (\alpha_i + \alpha_i^*) \end{cases} \quad (2)$$

Subject to

$$\sum_{i=1}^{n} (\alpha_i + \alpha_i^*) = 0,\ \alpha_i,\ \alpha_i^* \in [0, C] \quad (3)$$

This allows us to solve for w, as seen in Eq. (4)

$$w = \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) x_i, f(x) = \sum_{i=1}^{n} (\alpha_i + \alpha_i^*)\langle x_i,\ x \rangle + b. \quad (4)$$

This is called the support vector expansion, which means w can be completely described as a linear combination of the training data points. In addition, we do not need all x_i to describe w, hence, w is called the support vector.

Finally, the inner product $\langle x_i,\ x \rangle$ is nonlinear, hence, for $f(x)$ to be computed, a preprocessing of the training input x_i is carried out by mapping it into some feature space, this is done using the kernel tricks. In this work, a radial basis kernel function is utilized as seen in Eqs.(5)-(6);:

$$f(x) = \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) k(x_i, x) + b, \quad (5)$$

$$k(x, x') = \exp\left(-\gamma \cdot \| x - x' \text{AptCommand2016;}^2 \right). \quad (6)$$

Where γ is the kernel coefficient which controls the smoothness of the decision boundary i.e. the performance of the SVR model. Optimal setting of γ is difficult in practice as a high value result in overfitting while setting it too low degrades the performance. Hence, in this paper, grid search and cross-validation were used to determine the optimal value of γ.

Overall, SVR is a type of ML model that aims to predict a continuous outcome by finding the best fit line within a predefined margin of error, using support vectors and kernel functions to manage both linear and

non-linear data. It can accurately predict the output of a biomass experiment (within a small error margin ε) given few data points by using a comparatively low complexity model. It achieved this by simultaneously minimizing the loss of predicting each data point while reducing the model complexity. SVR like SVM is desired for its flexibility and robustness to outlier due the construction of the hyperplane margin of Eq. (2) which handles the trade-off between the decision boundary and error margin. Lastly, SVR easily finds a nonlinear decision boundary by mapping the data into higher-dimensional space using the kernel trick (function) as shown in Eq. (5).

2.2.2: Decision tree (DT)

DT is widely used for both classification and regression tasks, although it was initially developed for classification purposes but was extended to support regression analysis. The tree is grown by applying all the data to the root node and subsequently, the branches are iteratively split until a single value terminal node is reached where no further splitting can be done, this terminal node corresponds to the output (prediction) of the decision tree.

Each interior node of the tree corresponds to one input variable and is split into child nodes based on the values of the input variable. Splitting is done by iteratively applying an attribute value test, which divides each interior node into subsequent nodes. The goal of DT is to iteratively partition the input space to achieve minimum error between the estimated output at the terminal node and the target output.

The estimated output at the terminal node is as seen in Eqs. (7):

$$\widehat{y}_i = \frac{\sum_{j\in n_i} y_j}{|n_i|} \tag{7}$$

Where n_i is the leaf node i and $|n_i|$ is the number of samples in the leaf node i. For classification problems, the popular splitting criterion for DT is entropy and the Gini index. In contrast, the splitting criterion for regression problem is an impurity measure given by the least square deviation:

$$d(n_i) = \sum_{j\in n_i} \left(y_j - \widehat{y}_i\right)^2 \tag{8}$$

Where $d(n_i)$ is the impurity measured at node i. Hence, the splitting criterion is as given by Eq. (9):

$$\Delta d = d(n_p) - \alpha d(n_l) - \beta d(n_r) \tag{9}$$

Here n_p is the parent node, n_l and n_r are the left and right child nodes of n_p, respectively. The proportion of data samples assigned to the left and right nodes are measured by α and β, respectively. Therefore, our objective is to maximize Δd to accurately fit the data.

In simpler terms, we chose DT due to the way it intuitively split the data using the features present in the dataset. This allows us to measure the important features that contribute to HHV. Hence, the training dataset was used to train the DT to perform the splitting more accurately, which contributed to the overall performance.

2.2.3: Random forest (RF)

RF is a non-parametric ensemble predictor (or classifier) that uses bootstrap aggregation (or bagging) technique to improve its prediction result. Given inputs $\boldsymbol{x}$ in the form of a random vector such that $\boldsymbol{x} \in \chi$, (where χ is the input distribution) and target output $y \in \mathbb{R}$ to form a training dataset: $\mathscr{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)\}$. The RF regressor randomly combines the prediction of many DTs $\tau(\boldsymbol{x}; \theta_k, \mathscr{D})$, $k = 1, 2, 3, \ldots, K$ where θ_k are the independent and identically distributed (idd) random variable vectors associated with each tree, this variable is used to sample the dataset for growing individual trees and to select the direction of splitting.

The RF regressor produces an unweighted average output from a finite number of DT predictors as shown in Eqs. (10) - (12).

$$\widehat{y}(\boldsymbol{x}; \theta, \mathscr{D}) = \frac{1}{K}\sum_{k=1}^{K} \tau(\boldsymbol{x}; \theta_k, \mathscr{D}) \tag{10}$$

RF improves the overall prediction of each DT predictor by minimizing the variance between the prediction of each tree $\tau(\boldsymbol{x}; \theta_k, \mathscr{D})$ which in turn avoids overfitting of the training examples.

$$\delta_m = \mathbb{E}[y - \widehat{y}]^2 = \frac{1}{K}\sum_{n\in\mathscr{D}} (y - \widehat{y}(\boldsymbol{x}; \theta, \mathscr{D}))^2 \tag{11}$$

According to the law of large numbers, as the number of trees grows infinitely, the variance tends to zero.

$$\delta_\infty = \lim_{K\to\infty} \mathbb{E}[y - \widehat{y}]^2 = \frac{1}{K}\sum_{n\in\mathscr{D}} (y - \widehat{y}(\boldsymbol{x}; \theta, \mathscr{D}))^2 \to 0 \tag{12}$$

Hence, RF estimates the target values by bootstrapping and aggregating the prediction of the DT predictors and minimizing the variance between their predictions. This way, RF prevents overfitting and generalizes well to unseen data.

The bagging technique of RF is more essential in this experiment: (i) it prevent model overfitting the limited dataset by training different DT on subsets of the data, (ii) reduces variance of the model by averaging their result and (iii) robustness to outliers.

2.2.4: Extreme gradient boosting (XGboost)

XGBoost is also an ensemble learner that uses boosting technique to build a robust estimator from weak predictors which are shallow DTs. XGboost combines the strength of these weak predictors to produce an accurate estimate of the target output. XGboost is desired for its flexibility, scalability, and explainability. For the purpose of this experiment, XGboost can identify and rank each feature (parameters) of the experiment, such that we can make a better prediction with few observable features.

Given the dataset $\mathscr{D}$ and K-number of trees $\tau(\boldsymbol{x}_i)$ as defined, XGboost additively constructs an estimate to make a prediction as shown in Eq. (13):

$$\widehat{y}_i = \sum_{k=1}^{K} \tau_k(x_i) \tag{13}$$

Unlike RF, XGboost is prone to overfitting, hence, a regularized objective function is constructed, which consists of a loss function and the regularization function as shown in Eq. (14)

$$\mathscr{J} = \sum_i \mathscr{L}(y_i, \widehat{y}_i) + \sum_k \Omega(\tau_k) \tag{14}$$

Where $\mathscr{L}(\cdot)$ is the loss incurred at each terminal node and $\Omega(\cdot)$ is the regularization function for the k^{th} tree. In this work, a least squared loss function given in (8) is used, while a L-2 norm ($\frac{1}{2}\lambda \parallel w\text{AptCommand2016;} \parallel^2$) is used as the regularization term.

In gradient boosting, (14) is iteratively approximated using a second-order approximation as shown in Eq. (15):

$$\mathscr{J}^{(t)} \simeq \sum_i \left[\mathscr{L}\left(y_i, \widehat{y}_i^{(t-1)}\right) + \nabla_{\hat{y}^{(t-1)}}\mathscr{L}\left(y_i, \widehat{y}_i^{(t-1)}\right)\tau_k(x_i) + \nabla^2_{\hat{y}^{(t-1)}}\mathscr{L}\left(y_i, \widehat{y}_i^{(t-1)}\right)\tau_k(x_i)\right] + \Omega(\tau_k) \tag{15}$$

Where ∇ and ∇^2 are the first and second-order derivatives of the loss function with respect to the prediction $\hat{y}$ at the previous iteration. By neglecting the constant in the derivatives, and substituting the objective function as shown in Eq. (16)

$$\begin{aligned}\overline{\mathscr{J}}^{(t)} &= \sum_i \left[\nabla_{\hat{y}^{(t-1)}}\mathscr{L}\left(y_i,\ \hat{y}_i^{(t-1)}\right)\tau_k(x_i) + \nabla^2_{\hat{y}^{(t-1)}}\mathscr{L}\left(y_i,\ \hat{y}_i^{(t-1)}\right)\tau_k(x_i)\right] + \gamma T \\ &\quad + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2, \overline{\mathscr{J}}^{(t)} \\ &= \sum_j^T \left[\left(\sum_{i\in j}\varphi_i\right)w_j + \frac{1}{2}\left(\sum_{i\in j}\phi_i + \lambda\right)w_j^2\right] + \gamma T\end{aligned} \tag{16}$$

Compared to the other models, XGboost learns complex patterns in the dataset by sequentially building simple (weaker) models which focuses more on the portions misclassified by the previous models. Incremental learning can effectively find an advanced splitting criterion compared to DT and RF, and effectively handles missing values and outlier. In addition, the regularization term both penalizes complex models and reduces likelihood of overfitting. Lastly, XGboost offers a comprehensive order of feature importance given by Eq. (17).

$$F_{score}(j) = \frac{\sum_{i=1}^{T}\overline{\mathscr{J}}^{(i)} \times W_{split}(j,\ t_i)}{\sum_{i=1}^{T} W_{split}(j, t_i)} \tag{17}$$

Where: $F_{score}(j)$ is the estimated feature importance score for feature j, T is the total number of trees in the ensemble and t_i represents each tree. The weight of the split, $W_{split}(j,\ t_i)$ indicates how often a feature is used to split the data.

The F- score is metric that quantifies the importance or contribution of each feature to the predictive power of the model. Although the term "F-score" can sometimes refer to a statistical measure used in hypothesis testing (the F-test), when talking about feature importance in ML, it usually refers to something different. It is used to quantify the relative importance of each feature or input variables on the output. In a decision tree, an F-score for a feature can be calculated based on how useful that feature is at reducing uncertainty or impurity in the data. This is often measured using a criterion such as the Gini impurity or information gain/entropy (Divine et al.,2024). Each time a feature is used to split the data, the improvement in the splitting criterion is accumulated. This improvement is a reflection of how well the feature separates the data into classes for classification problems, or how well it splits the data into groups with different output values for regression problems. Moreover, in the ensemble methods like RF or GBR, each tree in the ensemble will have its own measure of feature importance. The overall F-score for a feature is usually the average across all trees in the ensemble, reflecting the average improvement in splitting criterion brought by that feature. In some implementations, it can also be the sum of these improvements.

2.3. Model set up

The study was conducted using a series of programs written in Python within the Google Colab environment. The key sequential methodology used in data preprocessing is shown in Fig. 2 and explained herewith.

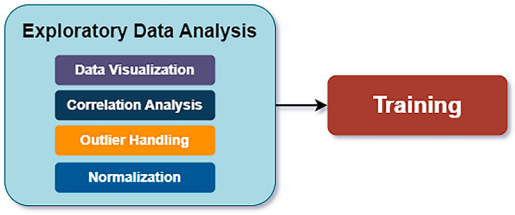

Fig. 2. Methodology used in exploratory data analysis.

Exploratory Data Analysis (EDA): We start the experiment by carrying out EDA on the dataset. This analysis aims to observe the relationship that exists among the different parameters of the dataset. We used regression plots, joint probability distribution plots, and covariance matrix tools to observe these relationships. The regression plot shows parameters of the dataset with high predictive relevance while the covariance matrix evaluated the degree of this relevance. The joint probability distribution plot also reveals an underlying relationship between each parameter, and additionally, reveals the presence of outliers in the dataset.

Outlier Handling: Outliers are data points that fall outside of the expected range of values in a dataset. They can significantly affect our analyses, leading to skewed results and incorrect predictions. Therefore, it's important to handle outliers appropriately in data analysis. In this work, a z-score technique was employed. The technique involves the following steps: we calculate the mean and standard deviation of the dataset, and the z-score for each data point by subtracting the mean from the data point and dividing it by the standard deviation. Then any data points with z-scores greater than 3 or less than −3 are flagged as outliers and replaced by the means of that column. The z-score threshold value is based on the popularly known empirical rule that 99.7 % of a normally distributed data falls within standard deviation of 3 i.e. ±3 z-score value, and any data outside this range is an outlier.

Normalization: this process standardizes the dataset by scaling it between 0 and 1 or −1 and 1. The latter was used in his paper so that the data has a 0-mean value (center around the mean).

Training: the training was carried out using the Python sklearn package. Four machine learning models were trained, and their training process and parameters are discussed herewith.

To select the optimal model parameter, we employed an exhaustive grid search to locate the optimal combination of the model parameters that minimizes root-mean-squared error and mean average percentage error and maximizes r^2. The parameters used for each model are presented in Table 2 where $\#feature$ is the number of features in the dataset $\mathscr{D}$, and $Var(\mathscr{D})$ is the variance of the dataset. We then perform a five-fold cross-validation and report the performance. Using the optimal parameter for each model, we applied a test dataset on the model and reported the performance. These results are discussed in the next section.

Table 2
Model parameter search table.

Model	Parameter	Value	Optimal value
Decision Tree	Criterion	Friedman mean squared error (Friedman, 2001), squared error	Friedman mean squared error
	Maximum depth	2, 4, 8, 16, 32	4
Random Forest	Criterion	Squared error, absolute error, Friedman MSE, Poisson	Absolute error
	Maximum depth	2, 4, 8, 16, 32	4
Support vector machine	Kernel	Polynomial, radial basis function	Polynomial
	Degree	1, 2, 3, 4, 5	3
	Gamma values	[0.1, 1]	0.488
	C (regularization)	10, 50, 100, 120	100
XG Boost	Maximum depth	3, 4, 5, 10, 20	5
	Learning rate	0.005, 0.01, 0.015, 0.05, 0.1	0.05
	Regularization alpha	0.01, 0.001, 0.0001	0.001
	Regularization lambda	0.0015, 0.00035, 0.00045	0.0015

3. Results and discussion

3.1. Pearson correlation analysis

Pearson Correlation Analysis is a statistical method used to measure the strength of the linear relationship between two continuous variables. In the context of ML model development, its significance is multifaceted: It can be used to identify and eliminate redundant features. If two features are highly correlated, they may convey similar information, and one can be removed to reduce dimensionality without a substantial loss of information. This can help in simplifying the model and potentially improving performance (Divine et al., 2024). Furthermore, identifying correlations between variables, one can gain insights into the underlying relationships within the data. This can help in understanding how different features may affect the outcome and in constructing more interpretable models. Pearson Correlation helps in diagnosing multicollinearity, which, if present, can lead to unreliable and unstable estimates of feature importance or regression coefficients. The use of Pearson Correlation to understand the relationships between variables can guide data preprocessing steps to determine if additional information or data process is required. For instance, when features are highly correlated, Principal Component Analysis (PCA) can be used to transform the features into a set of uncorrelated variables, which can then be used in model training.

Pearson Correlation was implemented in this study to understand the underlying distribution of the dataset. The result of the Pearson correlation analysis is presented in Fig. 3 as a correlation matrix showing the coefficients of correlation. The correlation matrix shows the strength of the dependence of the parameters on each other and the direction of such a relationship. A negative coefficient shows an inverse relationship while a positive coefficient shows a direct relationship. As the values of the coefficient tend to ±1, the relationship between the parameters is strengthened.

In the heatmap, strong positive correlations are likely represented by darker shades of one color (perhaps purple), and strong negative correlations by another (perhaps orange), with neutral correlations in a lighter shade.

Carbon content (C%) and hydrogen content (H%) show a positive correlation with HHV, as seen by the coefficients of 0.573090 and 0.080030, respectively, suggesting that higher carbon and hydrogen contents are associated with higher energy content in the biomass. Volatile matter (VM wt%) and fixed carbon (FCa wt%) have negative correlations with ash content, implying that as the proportion of VM and FCa increases, the proportion of ash tends to decrease. The negative correlation of oxygen (Oa%) with HHV (−0.595983) suggests that higher oxygen content is associated with a lower HHV. These correlations align with the expectations that combustible elements (C, H) contribute positively to the energy value of biomass, while non-combustible constituents (ash, O) contribute negatively.

The further explore the relationship between the input variable and HHV of biomass a parity plot of C content and HHV values was presented in Fig. 4. The figure shows that there is a slightly positive relationship between C content and the HHV of the biomass. It should be mentioned that the results from the parity plot correlates with those from the heatmap.

To clearly show the outlier of each parameter, a probability distribution plot is employed, as shown in Fig. 5(a). Plot 5(a) features an isolated contour, which is separate from the main concentration of data, indicating the presence of an outlier. It should be mentioned that the outlier could potentially skew the analysis, leading to inaccurate correlations or models if not addressed appropriately. Apart from showing the outlier, the contour plot also shows the perfect underlying distribution of the dataset and what will happen when the outlier is taken away. Knowing there's an outlier in the dataset, the z-score algorithm was used to remove and replace the outlier. After this process, a clean dataset is obtained. After the outlier in Fig. 5(a) is replaced, the new distribution plot is shown in Fig. 5(b). As shown in Fig. 5(b), there is a noticeable absence of teh isolated contour, implying that the outlier has been removed or otherwise dealt with. As a result, the remaining data shows a more cohesive pattern, which may provide a more accurate reflection of the underlying relationship between carbon content and HHV in the biomass without the distortion caused by outliers. The removal of outliers helps in better understanding the true central tendencies and variations within the data, leading to more reliable statistical or ML models.

After outlier removal and all other preprocessing steps, the data is divided into training and test datasets using a ratio of 80:20 for model development and model prediction ability assessment.

3.2. Model prediction accuracy

Four models were trained using the dataset to predict the highest heating values of biomass. A five-fold cross-validation was performed on each model and the average result was reported in Table 2. Table 2 shows the cross-validation results in terms of root-mean-squared error (RMSE), mean average percentage error (MAPE), and R^2, along with their standard deviation.

In Table 2, higher values of R^2 imply the model have higher prediction power while lower values of RMSE and MAPE show that the model's prediction is not far from the actual values. In addition, the lower standard deviation on the cross-validation shows that the model does not overfit the training data. From Table 2, XG Boost performed best on the cross-validation across all evaluation metrics with a predicting power of 98.61 % as the R^2 value and lower error rates of 0.1061 (±0.209) and 0.0046 (±0.0091) for RMSE and MAPE, respectively. Compared with the RF, both the DT and SVM apparently performed worse. Hence, the higher standard deviation of a RF shows that it overfits the training data, although the extent of overfitting is minimal compared to a DT.

Further, we perform a grid search on a range of parameters for each

	VM (wt%)	Ash (wt%)	FCa (wt%)	C (%)	H (%)	Oa (%)	HHVexp (MJ/kg)
VM (wt%)	1.000000	-0.597585	-0.714938	0.129115	0.204499	0.034522	0.219228
Ash (wt%)	-0.597585	1.000000	-0.111922	-0.354672	0.027548	0.220968	-0.524612
FCa (wt%)	-0.714938	-0.111922	1.000000	0.126167	-0.287068	-0.218923	0.155458
C (%)	0.129115	-0.354672	0.126167	1.000000	0.055646	-0.878037	0.573090
H (%)	0.204499	0.027548	-0.287068	0.055646	1.000000	-0.167765	0.080030
Oa (%)	0.034522	0.220968	-0.218923	-0.878037	-0.167765	1.000000	-0.595983
HHVexp (MJ/kg)	0.219228	-0.524612	0.155458	0.573090	0.080030	-0.595983	1.000000

Fig. 3. Heatmap showing the correlation matrix of the dataset.

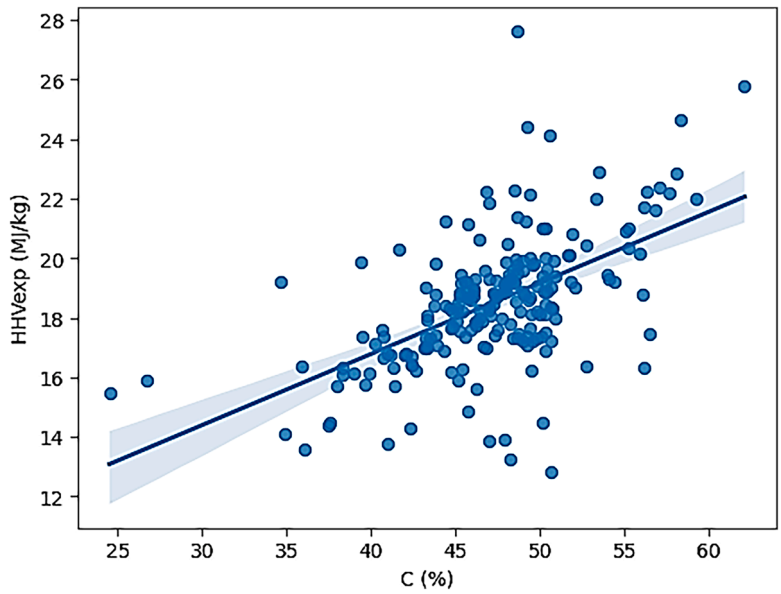

Fig. 4. Parity plot showing a slightly positive relationship between the C content and biomass HHV.

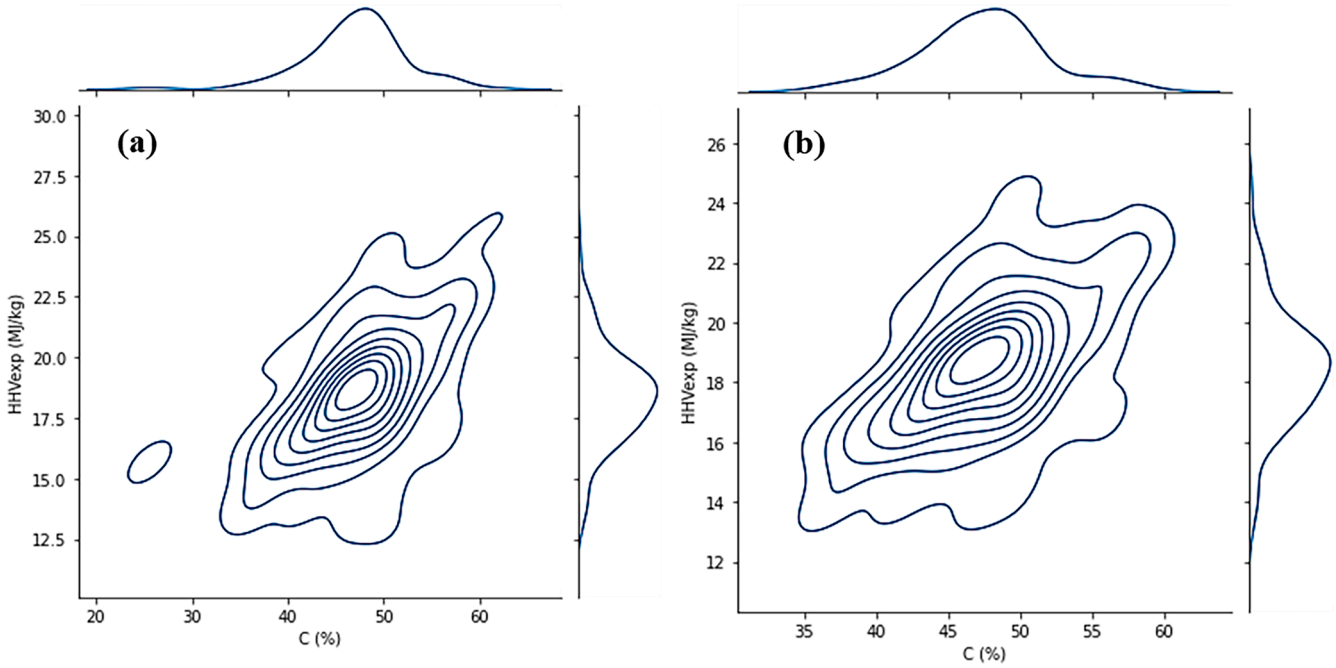

Fig. 5. (a) Distribution plot of the carbon content and HHV showing the presence of outlier, (b) Distribution plot of the carbon content and HHV showing the absence of an outlier.

model and select the best model using the four-fold cross-validation and the results are presented in Table 3. The best model with the optimal combination of these parameters was then used to train the model using the training dataset and evaluated using the test dataset. Table 3 shows the training and test results of each optimal model. It is evident that the XG Boost performs very well on both training and test datasets with R^2 values of 0.9683 and 0.7309, respectively; followed by SVM with R^2 values of 0.8642 and 0.6738, respectively. The performance of the DT on the training and test datasets was erratic, as the DT performed better on the test dataset and poorly on the training dataset.

Comparing the initial model results (Table 2) with those after hyperparameter tuning via grid search (Table 3), we observe changes in performance across all models and metrics. Notably, the RF and SVM models saw increases in both RMSE and MAPE in the test set, indicating a reduction in predictive accuracy and precision. The DT model exhibited an improved R^2 on the test set, which suggests a better fit to the variance of the data despite a large increase in test MAPE, hinting at possible overfitting issues. The most striking difference was observed with the XG Boost model; it underwent a considerable increase in test RMSE and a decrease in test R^2, moving from near-perfect to moderate explanatory power, although it still maintained the highest R^2 among all models post-tuning. This might suggest that the hyperparameter tuning for XG Boost was not optimal or that the model became more overfitted to the training data, losing its generalizability to new data. Overall, while tuning generally aims to improve model performance by optimizing the parameters, in this case, it seems to have resulted in trade-

Table 3
Comparative performance of machine learning models for HHV Prediction with hyperparameter optimization.

Model	RMSE		MAPE		R^2	
	Training	Test	Training	Test	Training	Test
Random Forest	1.1216	1.3184	0.0428	0.0534	0.6855	0.6827
Decision Tree	1.1833	1.0290	0.0474	0.4137	0.6499	0.8066
Support Vector Machine (SVM)	0.7369	1.3366	0.0195	0.0591	0.8642	0.6738
XG Boost	0.3558	1.2141	0.0148	0.0501	0.9683	0.7309

Table 4
Statistical criteria comparison between ML and empirical models. *Values calculated by linear extrapolation.

Model Type	Input features	RMSE	References
Empirical	Ultimate analysis	0.506*	Demirbas (1997)
Empirical	Proximate analysis	0.408*	Yin (2011)
ML (ANN)	Ultimate analysis	3.87	Xing et al. (2019)
ML (ANN)	Proximate analysis of biochars	0.65	Çakman et al. (2021
ML (RF)	Proximate and ultimate analysis, biomass classes	1.37	Afolabi et al. (2022)
ML (XGBoost)	Proximate and ultimate Analysis	0.1061	This study

offs between different measures of performance rather than clear improvements across the board.

The developed ML model in this study were compared with empirical models and other ML models reported in literature. The results are presented in Table 4. Based on the lower RMSE values of our proposed ML model, it can be deduced that the model performed satisfactory compared to empirical models and other ML models.

3.3. Feature evaluation

In the determination of the quality of fuel (biomass), a significant parameter to be considered is HHV (Boumanchar et al., 2019). Studies have utilized the ML model in the prediction of HHV of different biomass material types using either ultimate or proximate analysis as input parameters. There are divided opinions about whether the ultimate analysis or proximate analysis produces satisfactory HHV results when used as input parameters (Choi et al., 2014; Hosseinpour et al., 2017; Ighalo et al., 2020; Taki and Rohani, 2022). Therefore, the utilization of both ultimate and proximate analyses as input parameters for the ML model of HHV prediction is important. Hence, the contribution and relative importance of each input parameter is to be assessed. The influence of each input parameter on the HHV of biomass, which is the output feature, is necessary to be analyzed and observed. To do this, the XG boost model was explored to see the importance assigned to each feature. It should be mentioned that XG boost was selected before it is one of the most promising ML model based on the results in Table 3.The feature importance is visualized in Fig. 6. Each time a feature is used for the decision to split a tree, a feature score is assigned to it, and this score increases by the number of times a feature is considered before splitting the tree (see Eq. (17)). Hence, the higher the feature score, the more important the feature.

Several studies have reported the relationship between HHV and ultimate analysis, proximate analysis, and elemental composition parameters, respectively (Boumanchar et al., 2019; Sheng and Azevedo, 2005; Afolabi et al., 2022). From Fig. 6, carbon is most important with a weighted score of 0.2327, followed by VM with 0.2201 and Ash with 0.1985. H, FCa, and Oa contribute the least each weighted of 0.1316, 0.1150, and 0.1019, respectively. In fact, C is 56.2 % more important than Oa, while a combination of C, VM, Ash, and H contributes a weighted 0.7829 importance which means, they are responsible for 78.29 % prediction power of the models.

During combustion, the ash content significantly influences the transfer of heat and oxygen diffusion on the fuel (biomass) surface (Ajimotokan et al., 2019). For proper combustion, excessive ash content of biomass is detrimental to the HHV. An increase in the ash content of biomass results in decreased HHV (Sheng and Azevedo, 2005). The ash does not contribute to the released heat during the combustion of biomass. Hence, its presence should be minimal. The study of Afolabi et al. (2022) revealed ash as the most significant input feature that affects HHV prediction of biomass using the RF model. The reason for the influence of carbon being the foremost in this present study can be linked to the fact that HHV (calorific value) has a strong dependence on carbon (Boumanchar et al., 2019). The presence of higher C and H

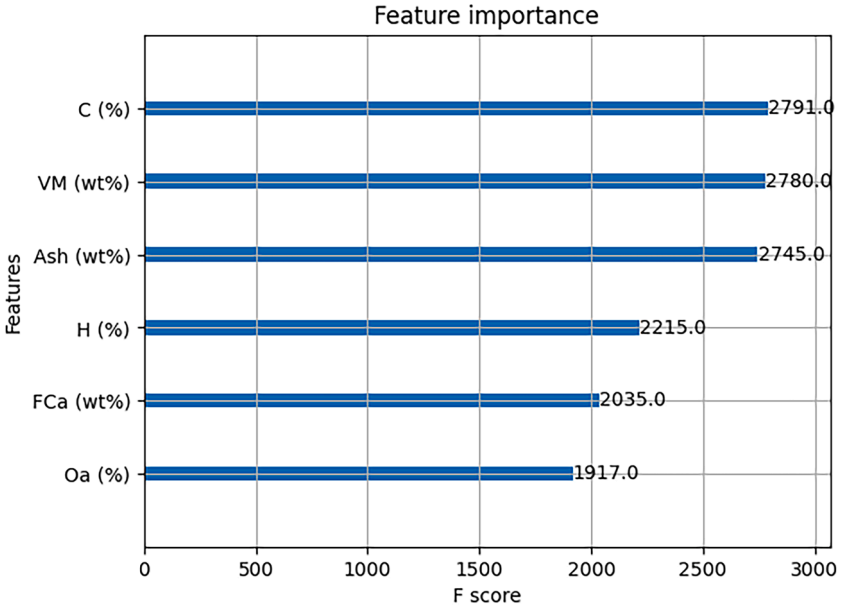

Fig. 6. Feature importance plot showing the relative relevance of each input variable on the output.

implies a higher energy content of biomass (Sheng and Azevedo, 2005; Balogun et al., 2021a; 2021b). It has been reported that there is a strong and positive correlation between HHV (output) and C and H, while the output has no significant correlation with oxygen. This is owing to the non-reactiveness of oxygen; hence, increasing oxygen content leads to decreased HHV (Sheng and Azevedo, 2005). In a similar work by Afolabi et al. (2022) where HHV was predicted using three different ML models, it was reported that ash, C, Vm, N, and biomass class were ranked consecutively as the input factors that contribute significantly to the prediction of the biomass HHV. It should be noted that the C content in biomass generally contributes to a higher HHV because carbon has a high energy content when burned. Thermodynamically, the combustion of carbon releases a significant amount of energy, typically in the form of heat, thus contributing to the HHV. Furthermore, from a thermodynamic perspective, the combustion of hydrogen produces a large amount of heat and water vapor, increasing the HHV.

4. Conclusion and future recommendations

In this study, the HHV of biomass materials was predicted using data-driven ML models with the incorporation of both proximate and ultimate analysis data. Four different models of RF, DT, SVM, and XG Boost were considered, which were best suited for non-linear and non-parametric problems. The XG Boost model was adjudged as the most accurate ML model due to its highest R-squared (R^2) values for both the training datasets (0.9683) and test datasets (0.7309), and lowest root mean squared error (RMSE) of 0.3558. The input parameters from both the proximate analysis and ultimate analysis revealed some level of relative significance based on the feature analysis done. Carbon (C), volatile matter (Vm), ash, and hydrogen (H) were considered the top four input features that significantly contributed to the prediction of the HHV. Hence, ML models are better suited for the prediction of HHV of biomass materials; thereby, reducing the cost and bottlenecks involved in using HHV experimental devices. This study further recommends that the HHV of torrefied biomass, hybridized torrefied biomass and coal, and biochar should be predicted using different ML models.

One notable limitation of the present study is the reliance on a relatively small dataset, with only 200 data points being considered. This restriction could potentially impact the generalizability and robustness of the findings. To address this issue, future research should explore the application of advanced deep learning techniques, such as Generative Adversarial Networks (GAN) or Variational Autoencoders (VAE), to generate synthetic datasets. This approach could significantly enhance the accuracy of machine learning models by providing a more comprehensive and diverse range of training data.

Another critical area for future exploration involves extending the prediction of higher heating value (HHV) to encompass a variety of carbonized materials, including coal, torrefied biomass, biochar, and hydrochars. It is essential for the developed machine learning model to effectively differentiate among these materials, as they each possess unique characteristics that can influence HHV. In the current study, the model primarily utilized proximate and ultimate analysis as input variables. However, to improve the precision and reliability of HHV predictions, subsequent studies should incorporate a more detailed compositional analysis. This includes examining the contents of cellulose, hemicellulose, and lignin, as well as understanding their respective impacts on HHV. Such an expansion of the analytical framework would not only refine the model's predictive capabilities but also deepen our understanding of the factors influencing the energy potential of various biomass materials.

CRediT authorship contribution statement

Adekunle A. Adeleke: Resources, Project administration, Methodology, Conceptualization. **Adeyinka Adedigba:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Steve A. Adeshina:** Methodology, Investigation, Formal analysis, Data curation. **Peter P. Ikubanni:** Investigation, Formal analysis, Data curation. **Mohammed S. Lawal:** Investigation, Formal analysis, Data curation. **Adebayo I. Olosho:** Writing – original draft, Formal analysis, Data curation. **Halima S. Yakubu:** Resources, Methodology, Formal analysis, Data curation. **Temitayo S. Ogedengbe:** Methodology, Investigation, Formal analysis, Data curation. **Petrus Nzerem:** Investigation, Formal analysis, Data curation. **Jude A. Okolie:** Writing – review & editing, Supervision, Software, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization.

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.

References

Adeleke, A.A., Ikubanni, P.P., Orhadahwe, T.A., Christopher, C.T., Akano, J.M., Agboola, O.O., Ibikunle, R.A., 2021. Sustainability of multifaceted usage of biomass: a review. Heliyon 7 (9).

Adeleke, A.A., Odusote, J.K., Ikubanni, P.P., Lasode, O.A., Malathi, M., Paswan, D., 2020. The ignitability, fuel ratio and ash fusion temperatures of torrefied woody biomass. Heliyon 6 (3).

Adeleke, A.A., Odusote, J.K., Lasode, O.A., Ikubanni, P.P., Malathi, M., Paswan, D., 2019. Densification of coal fines and mildly torrefied biomass into composite fuel using different organic binders. Heliyon 5 (7).

Adeleke, A.A., Okolie, J.A., Ogbaga, C.C., Ikubanni, P.P., Okoye, P.U., 2023. Machine learning model for the evaluation of biomethane potential based on the biochemical composition of biomass. Bioenergy Res. https://doi.org/10.1007/s12155-023-10681-9.

Adeshina, S.A., Adedigba, A.P., 2022. Bag of Tricks for improving deep learning performance on multimodal image classification. Bioeng. 9 (7), 312.

Afolabi, I.C., Epelle, E.I., Gunes, B., Güleç, F., Okolie, J.A., 2022. Data-driven machine learning approach for predicting the higher heating value of different biomass classes. Clean Technol 4, 1227–1241. https://doi.org/10.3390/cleantechnol4040075.

Ajimotokan, H.A., Ehindero, A.O., Ajao, K.S., Adeleke, A.A., Ikubanni, P.P., Shuaib-Babata, Y.L., 2019. Combustion characteristics of fuel briquettes made from charcoal particles and sawdust agglomerates. Sci. Afr. 6, e00202.

Al-Janabi, S., Mohammad, M., Al-Sultan, A., 2020. A new method for prediction of air pollution based on intelligent computation. Soft Comput 24 (1), 661–680.

Ayeleru, O.O., Dlova, S., Akinribide, O.J., Ntuli, F., Kupolati, W.K., Marina, P.F., Olubambi, P.A., 2020. Challenges of plastic waste generation and management in sub-Saharan Africa: a review. Waste Manag 110, 24–42.

Balogun, A.O., Adeleke, A.A., Ikubanni, P.P., Adegoke, S.O., Alayat, A.M., McDonald, A. G., 2021. Kinetics modeling, thermodynamics, and thermal performance assessments of pyrolytic decomposition of *Moringa oleifera* husk and *Delonix regia* pod. Sci. Rep. 11 (13862), 1–12. https://doi.org/10.1038/s41598-021-93407-1.

Balogun, A.O., Adeleke, A.A., Ikubanni, P.P., Adegoke, S.O., Alayat, A.M., McDonald, A. G., 2021b. Thermal decomposition and kinetic modeling of a tropical grass (*Digitaria sanguinalis*) under nitrogen and air environments. Case Stud. Therm. Eng. 26, 1–15.

Boumanchar, I., Charafeddine, K., Chhiti, Y., Boumanchar, I., Charafeddine, K., Chhiti, Y., M'hamdi Alaoui, F.E., 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. 9, 499–509. https://doi.org/10.1007/s13399-019-00386-5.

Çakman, G., Gheni, S., Ceylan, S., 2021. Prediction of higher heating value of biochars using proximate analysis by artificial neural network. Biomass Conversion and Biorefinery 1–9.

Choi, H.L., Sudiarto, S.I.A., Renggaman, A., 2014. Prediction of livestock manure and mixture higher heating value based on fundamental analysis. Fuel 116, 772–780.

Dai, Z., Chen, Z., Selmi, A., Jermsittiparsert, K., Denić, N.M., Nešić, Z., 2021. Machine learning prediction of higher heating value of biomass. Biomass Convers. Biorefin. 1, 1–9.

Demirbaş, A., 1997. Calculation of higher heating values of biomass fuels. Fuel 76 (5), 431–434.

... & Divine, D.C., Hubert, S., Epelle, E.I., Ojo, A.U., Adeleke, A.A., Ogbaga, C.C., Okolie, J.A., 2024. Enhancing biomass Pyrolysis: predictive insights from process simulation integrated with interpretable Machine learning models. Fuel 366, 131346.

Doddapaneni, T.R.K.C., Kikas, T., 2023. Advanced applications of torrefied biomass: a perspective view. Energies 16 (4), 1635.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics, pp. 1189–1232.

Güleç, F., Pekaslan, D., Williams, O., Lester, E., 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.

Hosseinpour, S., Aghbashlo, M., Tabatabaei, M., Mehrpooya, M., 2017. Estimation of biomass higher heating value (HHV) based on proximate analysis by using iterative neural netwook-adapted partial least squares (INNPLS). Energy 138, 473–479.
Ighalo, J.O., Adeniyi, A.G., Marques, G., 2020. Application of linear regression algorithm and stochastic gradient descent in a machine learning environment for predicting biomass higher heating value. Biofuels Bioprod. Biorefin. 14, 1286–1295.
Ighalo, J.O., Igwegbe, C.A., Adeniyi, A.G., 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, 3177–3191. https://doi.org/10.1007/s40808-021-01276-4.
Odusote, J.K., Adeleke, A.A., Lasode, O.A., Malathi, M., Paswan, D., 2019. Thermal and compositional properties of treated Tectona grandis. Biomass Convers. Biorefin. 9, 511–519.
Onsree, T., Tippayawong, N., 2021. Machine learning application to predict yields of solid products from biomass torrefaction. Renew. Energy 167, 425–432.
Scarlat, N., Fahl, F., Dallemand, J.F., 2019. Status and opportunities for energy recovery from municipal solid waste in Europe. Waste Biomass Valor 10 (9), 2425–2444.
Sheng, C., Azevedo, J.L.T., 2005. Estimating the higher heating value of biomass fuels from basic analysis data. Biomass Bioenergy 28 (5), 499–507.
Strubell, E., Ganesh, A., McCallum, A., 2020. Energy and policy considerations for modern deep learning research. Proceed. AAAI conf. Artificial Intellig. 34 (9), 13693–13696.
Taki, M., Rohani, A.A., 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.
Umenweke, G.C., Afolabi, I.C., Epelle, E.I., Okolie, J.A., 2022. Machine learning methods for modeling conventional and hydrothermal gasification of waste biomass: a review. Bioresource Technol. Reports 17, 100976.
Vallejo, F., Díaz-Robles, L., Cubillos, F., Perez, A., 2020. Valorization of municipal solid waste using hydrothermal carbonization and gasification: a review. Chem. Eng. 81.
Verger, T., Azimov, U., Adeniyi, O., 2022. Biomass-based fuel blends as an alternative for the future heavy-duty transport: a review. Renew. Sustain. Energy Rev. 161, 112391.
Xing, J., Luo, K., Wang, H., Gao, Z., Fan, J., 2019. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. Energy 188, 116077.
Yan, M., Cui, J., Li, T., Feng, H., Hantoko, D., Kanchanatip, E., 2023a. Transformation and distribution of nitrogen and phosphorus in sewage sludge during supercritical water gasification. Fuel 332, 125918.
Yan, X., Li, J., Smith, A.R., Yang, D., Ma, T., Su, Y., Shao, J., 2023b. Evaluation of machine learning methods and multi-source remote sensing data combinations to construct forest above-ground biomass models. Int. J. Digit. Earth 16, 4471–4491.
Yin, C.Y., 2011. Prediction of higher heating values of biomass from proximate and ultimate analyses. Fuel 90 (3), 1128–1132.