Ain Shams Engineering Journal 16 (2025) 103209

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Optimizing biochar yield and composition prediction with ensemble machine learning models for sustainable production

Jingguo Gou [a,*], Ghayas Haider Sajid [b,*], Mohanad Muayad Sabri [c], Mohammed El-Meligy [d,e], Khalil El Hindi [f], Nashwan Adnan OTHMAN [g,h]

[a] School of Architecture and Surveying Engineering , Shanxi Datong University, Datong 037003,China
[b] Department of Environmental Sciences, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
[c] Peter the Great Saint Petersburg Polytechnic University, Saint Petersburg, Russia
[d] Jadara University Research Center, Jadara University, PO Box 733 Irbid, Jordan
[e] Applied Science Research Center, Applied Science Private University, Amman, Jordan
[f] Department of Computer Science, College of Computer & Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[g] Department of Computer Engineering, College of Engineering, Knowledge University, Erbil 44001, Iraq
[h] Department of Computer Engineering, Al-Kitab University, Altun Kupri, Iraq

ARTICLE INFO

Keywords:
Biochar
Yield
Composition
Pyrolysis
Data-driven modeling

ABSTRACT

Biochar production from organic waste can reduce fossil fuel reliance and combat climate change, but current models are computationally demanding and have limited accuracy. The study creates four machine learning models using multiple linear regression, decision trees, Adaboost regressors, and bagging regressors, trained on a dataset of pyrolysis tests. The results show that the data-driven models have significantly higher predictive accuracy than existing models, with an R^2 of up to 0.96. The Bagging Regressor (BR) demonstrated superior efficacy compared over the MLR, AR, and DT models across all eight output parameters, with R2 values of 0.94, 0.93, 0.93, 0.94, 0.95, 0.90, 0.92, and 0.96 for Biochar Yield, Fixed Carbon, Volatile Matter, Ash, and ultimate composition parameters (C, H, O, and N), respectively. The study developed a data-driven model to predict Biochar yield and compositions, enhancing production processes and promoting sustainable farming practices.

1. Introduction

As concerns about the impact of climate change and the global energy crisis both worsen, the development of alternative energy sources is becoming increasingly important [1]. A biomass of about 140 billion metric tons is produced each year globally, and inappropriate disposal of this material pollutes the environment [2]. The conversion of wastes into useful products appears to be an effective way to manage waste and promote sustainable resource usage [3]. Biochar can be produced using thermochemical techniques from organic waste, an important type of waste biomass. Biochar is being known as one of the most valuable renewable bioresources due to its numerous applications, which include contaminant adsorption, GHG (greenhouse gas) elimination, treatment of wastewater, soil remediating, manufacturing of energy, and usage as catalysts. [4]. Surface functional categories, a sizable surface-specific area, and a high amount of minerals are some of its distinguishing features. Specifically, because of its carbon-rich microporous structure, it is used in a range of applications [5]. For the removal of water and air pollutants, biochar is a useful adsorbent. The outcome may be affected by the kind of feedstock utilized and the technique of production used, such as pyrolysis or gasification [6,7]. Biochar is used to absorb water pollutants and remove heavy metals from wastewater. According to scientific literature, the Successful removal of lead, Copper, Zinc, and Cadmium from water-based solutions was achieved by biochar made of husks of rice and manure as raw materials [8]. Furthermore, it has been demonstrated that biochar is effective at removing organic contaminants such as aromatic materials, organic solvents that are volatile and pungent chemicals, as well as metal aerosols, especially acid gases, ozone, metallic mercury (Hg), oxides of nitrogen, and other contaminants. [9]. Another important feature of biochar is its catalyst nature, which makes it suitable for an array of uses such as the syngas purification process, biofuels synthesis, and pollution in the atmosphere

* Corresponding authors.
E-mail addresses: jggousdu@163.com (J. Gou), ghayashaider1277@gmail.com (G.H. Sajid), mohanadsabri1@gmail.com (M.M. Sabri), mohammedali2000@gmail.com (M. El-Meligy), khindi@ksu.edu.sa (K.E. Hindi), nashwan.othman@knu.edu.iq (N.A. OTHMAN).

https://doi.org/10.1016/j.asej.2024.103209
Received 28 August 2023; Received in revised form 10 November 2024; Accepted 28 November 2024
Available online 28 December 2024

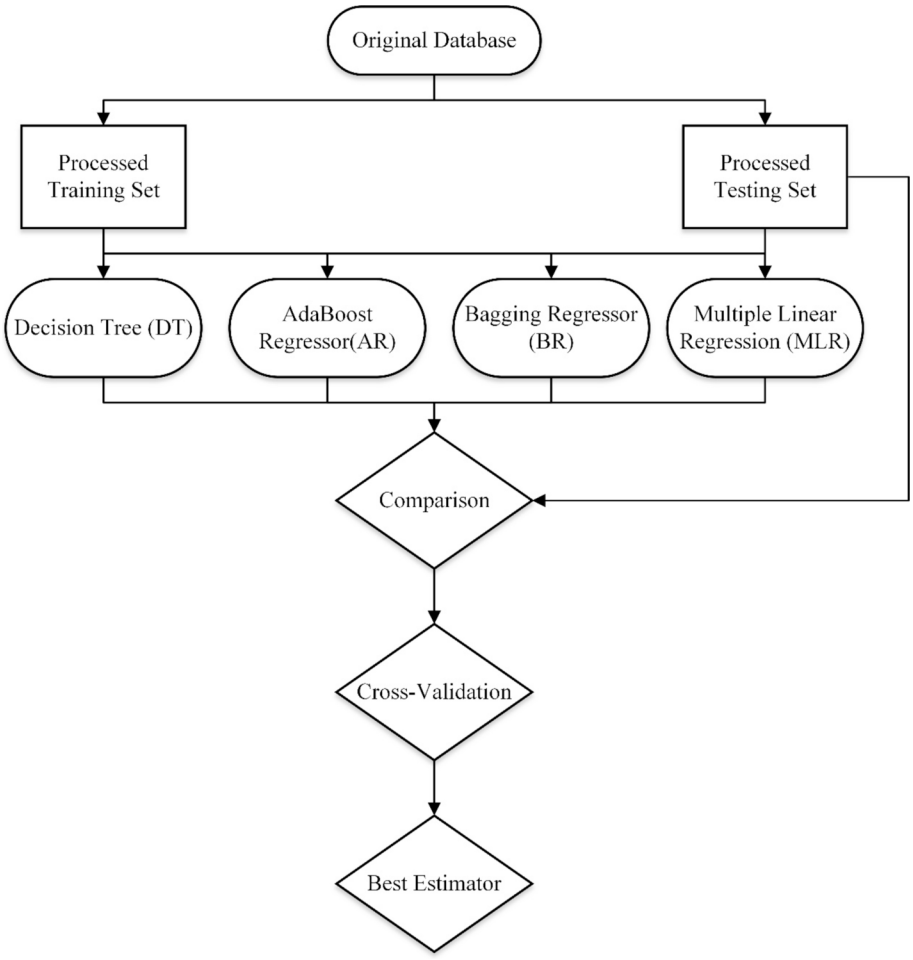

Fig. 1. Adapted process flowchart for this investigation.

Table 1
To ensure generalizability, a statistical analysis of the input and output factors for the raw data.

I/O	Types Of Variables	Variables	Mean	Median	Mode	Standard Deviation	Range	Minimum	Maximum	Count (Percent Missing)
Input	Feedstock Proximate composition	*Fixed carbon*	13.84	13.84	13.84	13.84	23.47	4.33	27.80	226(0 %)
		Volatile matter	79.83	80.85	82.38	4.91	22.96	68.20	91.16	226(0 %)
		Ash	6.33	6.61	7.45	3.94	14.98	0.16	15.14	226(0 %)
	Feedstock ultimate composition	*Carbon*	44.19	43.57	43.95	5.44	28.53	35.70	64.23	226(0 %)
		Hydrogen	5.94	5.81	5.81	1.00	6.08	4.10	10.18	226(0 %)
		Oxygen	42.44	42.00	41.12	5.20	25.49	27.61	53.10	210(16 %)
		Nitrogen	1.30	0.99	1.12	1.66	9.61	0.00	9.61	226(0 %)
		Sulphur	0.48	0.48	0.56	0.24	0.83	0.09	0.92	152(74 %)
	Feedstock lignocellulosic composition	*Cellulose*	37.52	40.14	28.70	8.24	29.78	17.89	47.67	112(114 %)
		Hemicellulose	24.97	23.21	39.30	13.30	44.81	11.48	56.29	107(119 %)
		Lignin	22.26	22.40	19.60	6.54	27.27	4.99	32.26	112(114 %)
	Pyrolysis condition	*Residence time (min)*	38.08	30.00	30.00	18.76	89.00	1.00	90.00	226(0 %)
		Temperature (°C)	460.66	450.00	600.00	124.82	600.00	200.00	800.00	226(0 %)
		Heating rate (°C/min)	11.37	10.00	10.00	5.69	20.00	5.00	25.00	226(0 %)
Output	Process Efficiency	*Biochar yield (%)*	39.53	35.20	23.30	15.10	78.21	17.68	95.89	226(0 %)
		HHV (MJ/kg)	23.28	22.50	20.56	5.50	34.06	3.60	37.66	105(121 %)
		Energy yield (%)	60.07	54.71	56.89	14.67	61.40	38.40	99.80	87(139 %)
	Biochar Proximate Composition	*Fixed Carbon*	53.57	54.16	59.20	19.77	79.07	15.04	94.11	159(67 %)
		Volatile matter	32.38	25.78	32.21	20.00	82.23	0.49	82.72	159(67 %)
		Ash	14.05	14.20	6.70	9.02	37.59	0.32	37.91	159(67 %)
	Biochar Ultimate Composition	*Carbon*	64.57	62.26	57.00	12.07	50.49	44.12	94.61	162(64 %)
		Hydrogen	3.69	3.59	2.91	1.50	7.46	1.26	8.72	150(76 %)
		Oxygen	17.39	14.38	9.64	10.27	45.17	0.00	45.17	150(76 %)
		Nitrogen	1.48	1.12	0.00	1.38	9.05	0.00	9.05	162(64 %)
			0.53	0.52*Sulphur*	0.47	0.31	1.29	0.00	1.29	115(111 %)

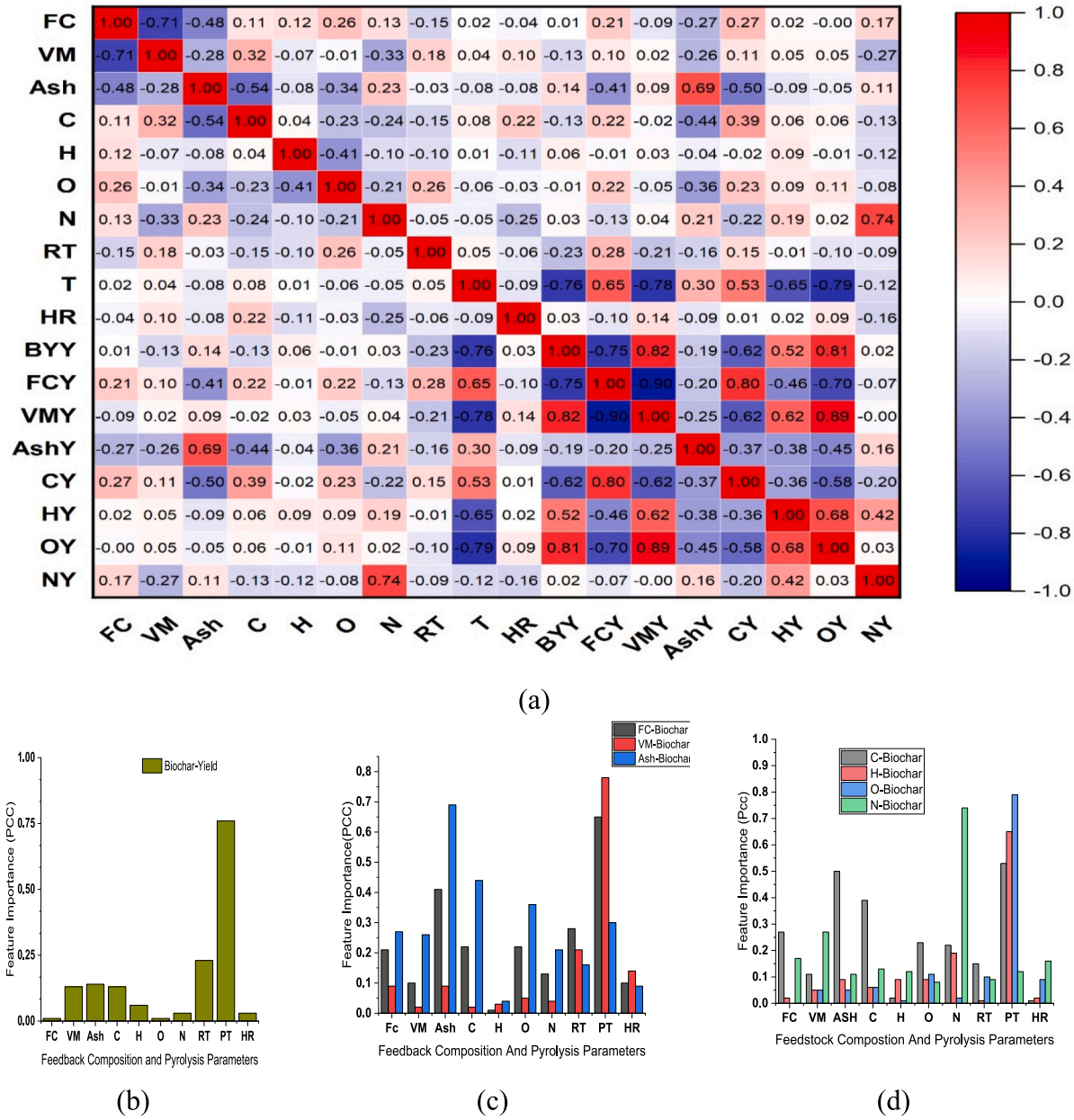

Fig. 2. (a) PCCC between each of the interested variables. b) Biochar yield; c) the approximate composition of biochar (FC-VM-ash); and d) the eventual composition of biochar (C—H—O—N).

reduction. [10]. Furthermore, biochar provides exceptional benefits for soil rehabilitation. When applied to soil, a part of the carbon in recalcitrant biochar might continue for a long time in the soil ecology and store carbon. [9]. According to a previous study, fertilizer made with decentralized, renewable energy sources emits fewer pollutants than fertilizer made with Conventional manufacturing methods [11].

One of the most often used processes for creating biochar is pyrolysis. Value-added products like biochar, bio-oil, and syngas are created when organic material is heated in an inert environment (i.e., without oxygen (O)) [10]. The pyrolysis process factors, in addition to the varieties and composition of the biomass source govern how applying biochar will affect the soil's condition and carbon reduction. For example, increased temperatures, greater combustion rates, reduced pressures, or smaller dimensions of particles during pyrolysis increase the breakdown of biomass polymers and reduce the generation of biochar [11]. Higher pyrolysis temperatures, on the contrary, result in a higher carbon content and, as a result, a higher quality biochar for soil usage. This conflict of interests demonstrates that the pyrolysis process produces biochar optimally under these circumstances [12]. In order to create the best biochar systems for ecologically friendly chemical and agricultural uses, it is essential to understand and forecast output.

Theoretical models have been frequently employed in the previous decades to forecast the product yields of pyrolysis processes [13]. Specifically, over the last two decades, a number of theoretical models have been employed extensively to forecast the yield of pyrolysis processes, including decision trees, support vector machines (SVM, neural networks, and integration methods for decision trees. For instance, with a model accuracy of R2 = 0.78–0.87 [12,13]. Li et al. employed feedback neural networks to forecast the generation of biochar from organic waste, demonstrating the impact of ideal model parameters determined by sensitivity [14]. Gradient learning outperformed random forests in Leng et al.'s machine learning prediction of biochar generation by

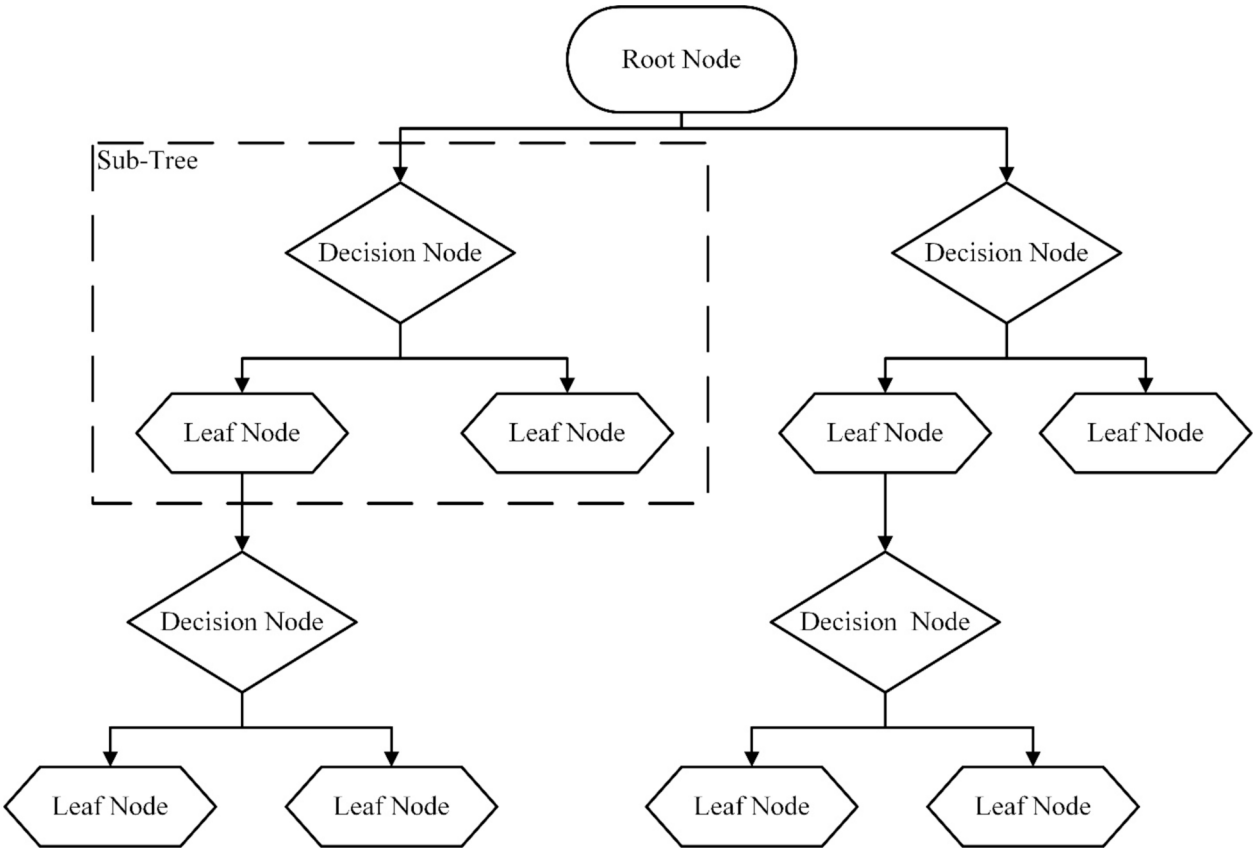

Fig. 3. Schematic Diagram of DT.

regulating carbonaceous ingredients [15].These models' semi-empirical character, complexity, and time requirements, however, limit their applicability for predicting biochar yield in favor of more Development and optimization of complicated processes (such as optimization with multiple goals of biochar application). Based on the results of the mathematical models or experiments, researchers have also created actual correlations [14]. However, because they are limited to a few specific experimental settings and biomass feedstocks, these empirical correlations are often not suitable for extrapolative situations.

Data-driven modelling has grown in favour as a tool for forecasting biochar synthesis, thanks to the advancement of artificial intelligence and the availability of an abundant amount of pyrolysis information from experiments. These techniques have better prediction accuracy, faster computation times, and capacities to reproduce complicated data trends [16]. The method in particular, using scant experimental and system data, could accurately estimate biochar production. Through training, it establishes the connection between the parameters affecting input and output. and generates conclusions that are unaffected by preconceived notions. However, current Machine learning (ML) forecast based biochar models for forecasting have a limited degree of accuracy, particularly when only a small number of datasets are used to build the model. In one earlier study, 245 datasets from a variety of biomass feedstocks and operational methods parameters were used to create a RF (Random Forest) regression-based model for estimating the output and carbon content of biochar. R2 values for the coefficients of determination are 0.855 & 0.848, respectively, were obtained by the researchers for projecting the output of biochar and the carbon content [15]. Another study looked at the accuracy of the xtreme gradient-boosting (XGB) machine learning be used to estimate biochar output. Using 91 data sets for training, its R2 value was 0.84 [16]. A recent research improved R2 for estimating biochar output by using an artificially generated neural network (ANN) in combination with metaheuristic approaches [17]. Popular prediction techniques include ANFIS (Adaptive Neuro-Fuzzy Inference System) and MLP-NN (Multiple Layer Perceptron Neural Networks), well-known based on data programming method for MLP-NN performs processing of signals, workflow training, functional estimating, and recognition of patterns. They are frequently employed to represent intricate relationships between input and output parameter spaces, examine data trends, or represent statistical sophistication in unidentified composite distributions of probabilities between variables that can be observed [18]. In contrast, ANFIS integrates adaptive control approaches with ANN & fuzzy inference techniques. Fuzzy logic is a technique for deep learning that allows human perception as well as decision-making uncertainty to be mathematically expressed [19]. The models' potential for predicting biochar output is still far from clear. However, the development of a comprehensive data-driven model capable of forecasting biochar output and compositions (both proximal and ultimate) has never been undertaken previously. According to the literature, pyrolysis conditions and biomass feedstock compositions can have a significant influence on biochar output and compositions. As a result, this study aims to build a comprehensive machine learning (ML) model capable of forecasting both biochar output and composition, which is of considerable scientific significance.

The previous approaches of estimating the yield and composition of biochar are limited by computational complexity and accuracy problems [17–20], which makes this research important. Such shortcomings highlight a crucial research gap and seriously impede the useful use of these techniques in industrial settings. More precise, dependable, and computationally efficient models that can handle the intricacies of the processes involved in the synthesis of biochar are desperately needed. While biochar yield has been predicted using machine learning models like Random Forest and Artificial Neural Networks (ANN), their forecast accuracy is usually modest, with R^2 values ranging from 0.84 to 0.855. The possibility of overfitting is another main problem with machine

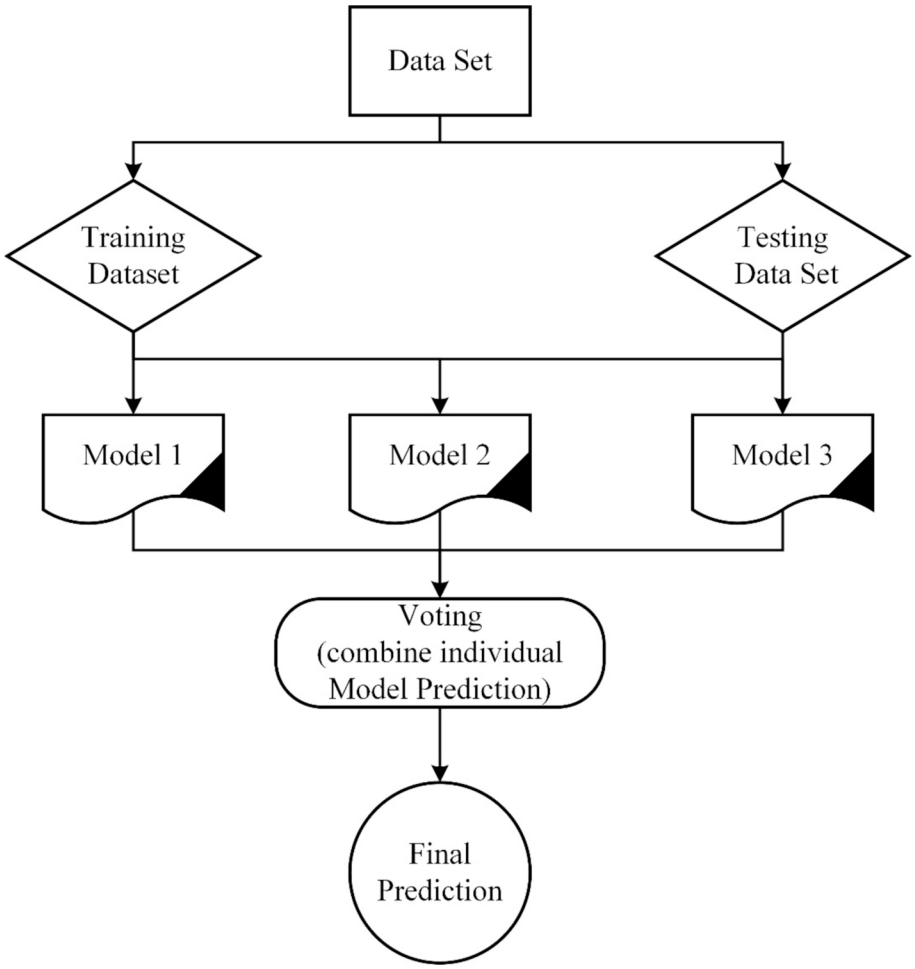

Fig. 4. shows the AdaBoost technique's schematic diagram.

learning models, especially when training datasets are small or insufficiently diverse. Because of this, models may perform well on training data but badly on unseen data, which could compromise their dependability in practical situations [19]. Many machine learning models function as "black boxes," which makes it challenging for users to comprehend how predictions are influenced by input characteristics. Because stakeholders must base their decisions on the projections, this lack of transparency may erode their confidence in the model's outcomes. Although ongoing, improving interpretability is still quite difficult [21].These constraint are especially noticeable when working with small datasets or a variety of biomass feedstocks. Furthermore, the models' high computational expense and scalability issues limit their practical application in optimizing the synthesis of biochar. This study stands out because it concentrates on improving accuracy by choosing input and output characteristics with greater diligence. To enhance data quality, we employ a broader range of biomass properties and sophisticated preprocessing methods. By addressing the shortcomings noted in earlier research, this methodological change could result in greater R2 values and more accurate biochar yield projections.

According to the author, this is a pilot study to develop a detailed model that, given the pyrolysis, would forecast biochar production and composition concurrently settings and compositions of the biomass feedstock. By focusing on multiple linear regression (MLR), decision trees (DT), Adaboot Regressor (AR), Bagging Regressor (BR) this study pioneers a comprehensive exploration of machine learning techniques for biochar yield and composition prediction. The compositions of different organic waste feedstocks (mostly straws and wood-based), associated pyrolysis factors, biochar productivity, and biochar compositions were all taken into account during the data assimilation step. Two methods, the RMSE (root mean square error) and R2, utilized to examine the effects of different model parameters and the division between training and testing datasets. A thorough performance comparison conducted between data-driven models and current models, which substantially produces a higher prediction accuracy, with an R2 value of up to 0.96. In addition, the research delves into how input parameters impact what is anticipated, offering an important comprehension of the factors influencing the production of biochar.

The goal of this work is to overcome these important constraints by creating a set of ensemble machine learning models that ensure computational efficiency while improving forecast accuracy. By closing this gap, the effort hopes to make it easier to use these sophisticated models in practical settings, which will ultimately lead to more efficient biochar production optimization and forward the field's quest for sustainable energy alternatives. The overall procedure for this inquiry is shown in Fig. 1. The recommended individual and ensemble approaches are described in the ensuing subsections.

2. Materials and methods

2.1. Data Collection and Pre-Processing

To develop a data-driven models, 226 datasets from the literature (19 studies) were acquired, is summarized in Table 1. The dataset contains corn cob, corn stalks, bagasse, coconut peat, coconut shells, coconut fibres, and other feedstocks. Wheat straws, husks of rice, straw from rice, pine trees, pine wood shavings, pine wood, bamboo, citrus

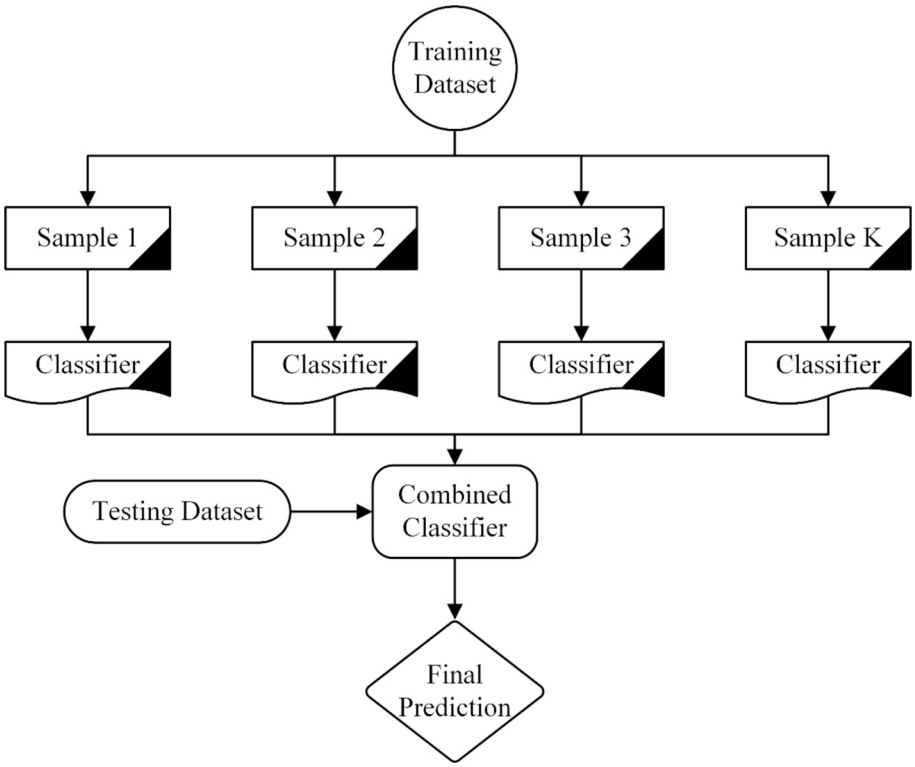

Fig. 5. Bagging Algorithm Schematic Diagram.

bagasse, orange pomace, and various other substances are among the materials used. Agricultural waste, The cassava plant roots, Cassava roots, Rape branch Wood stem, Wooden bark, Cotton stem Canola hull, Oats hull, Vines pruning, Hinoki cypress Litter of poultry [22–40].

To ensure the data's broad applicability and generalizability, the following aspects were considered throughout the data gathering stage: (1) feedstock biomass proximate composition, (2) ultimate biomass feedstock composition, (3) lignin-based biomass feedstock composition, (4) the primary pyrolysis circumstances, (5) the conditions for biochar manufacturing, (6) the near-ultimate component of the biochar, (7) the higher heating value (HHV) of the biochar, & (9) the overall energy yield of the biochar. Despite the fact that prior efforts have taken into account feedstock particle size [41,42]. Due to significant methodological variances and data-gathering process concerns, it was excluded from the current analysis.

Ash, fixed carbon, volatile matter, and biochar were the closest compositional correlations. It is important to note that the current information is restricted to feedstock ash percentage ranges of 0–15 %. For increased ash content input data, future dataset development will be required to create data-driven algorithms. Carbon, hydrogen, nitrogen, oxygen, and sulfur are the primary (or elemental) components of feedstock and biochar. For both feedstock and biochar, the literature had both wet-base and dry-based data, which were converted to dry material to conduct the task Using Eq. (1–3). MC stands for moisture content. Three elements make up the feedstock's lignocellulose composition: cellulose, hemicellulose, and lignin. The three most important pyrolysis process parameters—residence duration, decomposition temperature, and heating rate—were identified in the literature. The Supplementary Material contains a full copy of the dataset.

$$FC_{dry}\frac{\mathrm{FC}wet}{1-\mathrm{MC}} \tag{1}$$

$$VM_{dry}=\frac{VMwet}{1-MC} \tag{2}$$

$$Ash_{dry}\frac{ashwet}{1-MC} \tag{3}$$

Because the data was gathered from an array of literature (see [14] [cite original paper here] for more information), discrepancies in the datasets were unavoidable, resulting in missing values (see Table 1). Certain characteristics were eliminated if more than 20 % of the data for the input parametric space was not available. On the basis of this, the feedstock was excluded from model development due to its high sulfur level and composition of lignin and cellulose. The output dataset's cutoff criteria were set at 35 %, which resulted in the removal of Sulfur concentration in biochar, HHV, and energy output. The updated dataset for the model now includes ten input parameters related to biomass feedstock composition and pyrolysis conditions: Fixed Carbon (FC), Volatile Matter (VM), Ash content (Ash), Carbon (C), Hydrogen (H), Oxygen (O), Nitrogen (N), Reaction Time (RT), Temperature (T), and Heating Rate (HR). Additionally, it features eight output parameters that describe biochar production and composition: Biochar Yield (BYY), Fixed Carbon Yield (FCY), Volatile Matter Yield (VMY), Ash Yield (AshY), Carbon Yield [43], Hydrogen Yield (HY), Oxygen Yield (OY), and Nitrogen Yield (NY).

The restored data continued to have a large number of values that are missing, which could result in incorrect model training. This issue was avoided by replacing the value of the attribute's missing values with the feature's average [38], guaranteeing an uninterrupted data. The model's performance during testing is unaffected by this strategy because it is only applied during the training phase. Data normalization was carried out as a required pre-processing step because the dataset contains variables with a variety of values, mean, and SD (Standard Deviation), as shown in Table 1. The conventional normal variable Zi (Equation (4) was utilized, as is typical in machine learning, and is represented as below.

$$Z_i\frac{Xi-\rho}{\sigma} \tag{4}$$

Xi represents raw data, ρ represents mean, and standard deviation

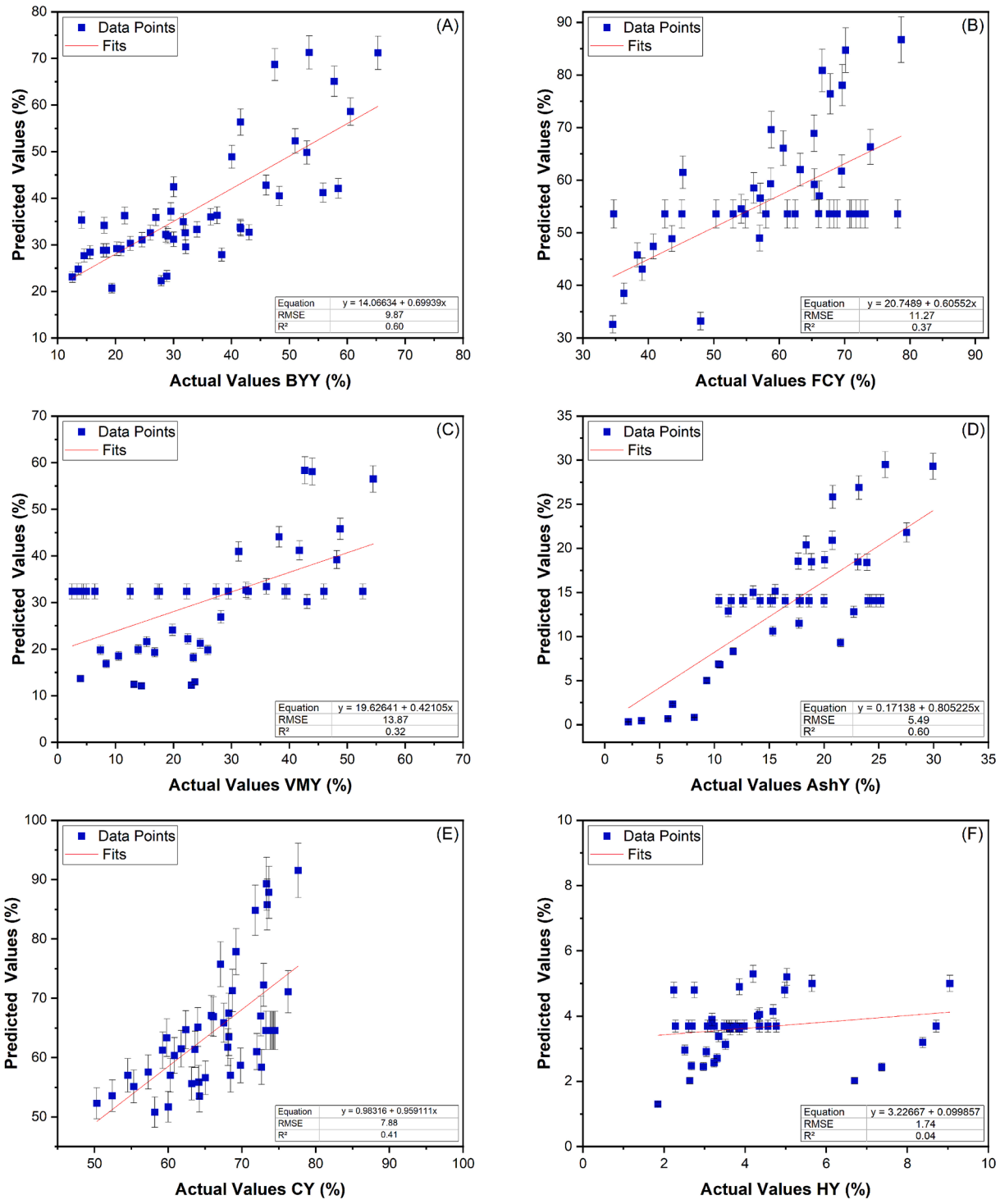

Fig. 6. Actual and predicted biochar yield, proximal composition (FC-VM-ash), and ultimate composition (C—H—O—N) are all plotted using multiple linear regression (MLR).

denoted by σ.

Referring to Eq. (5), a coefficient called the PCC (Pearson Correlation Coefficient) was utilized to calculate the degree of linear correlation between a pair of variables (i.e., inputs and outputs or two distinct inputs) [15]. A PCC of zero denotes no correlation, while a PCC of one indicates a significant correlation exists between the factors. The formation of biochar, proximal composition, and final composition were three output factors that this absolute PCC value also demonstrated the relative importance of. Section 2.2 discusses the PCC findings and the relative weights of different factors.

$$PCC = \frac{\sum_{i=1}^{n} Xi - \overline{X}\sum_{i=1}^{n}(y_i - y)}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (5)$$

In this case, the two important variables from which PCC must be calculated are x and y, and the number of data points is n.

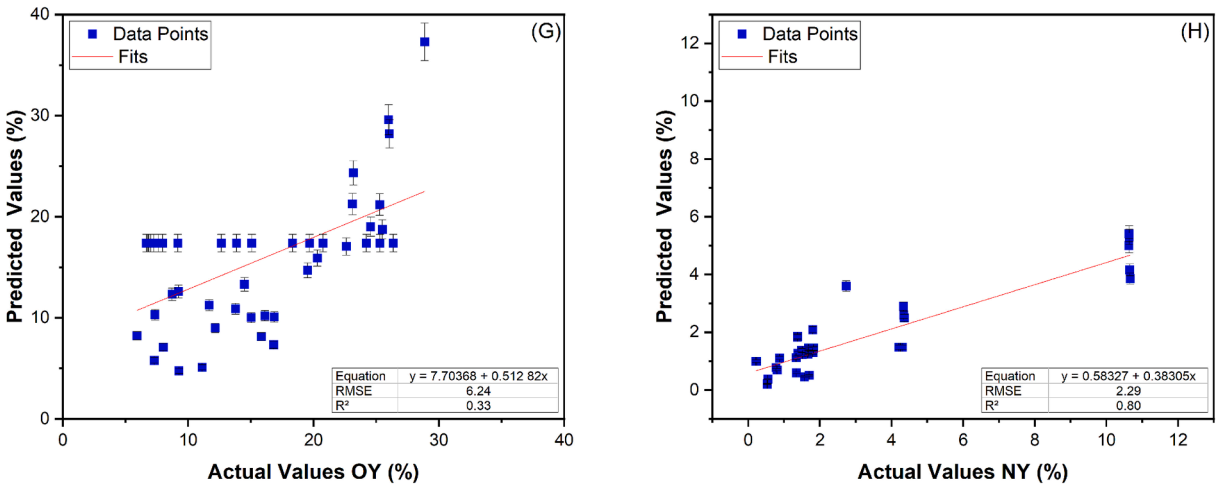

Fig. 6. (*continued*).

2.2. Exploration of dataset

Table 1 shows the statistical properties of the input parameters i.e FC, VM, Ash, C, H, O, N, RT, T, and HR and output parameters i.e BYY, FCY, VMY, AshY, CY, HY, OY, and NY. The PCC (Pearson correlation coefficient) was used to assess the linear correlation between the variables in question. A relationship is considered to be weak if the PCC value is close to 0, and strong if the value is close to 1.

The Fig. 2(a) shows that strong relationships (with PCC $>$ 0.35) exist between the input parameters and the different elements that make up the proximate and final composition of biomass feedstock. For example, there is a strong negative correlation between FC and VM (PCC $=$ -0.71), which means that as the FC content of the biomass feedstock increases, the VM content decreases. Likewise, there is a strong negative correlation between C and Ash (PCC $=$ -0.54), which means that as the C content of the biomass feedstock increases, the Ash content decreases. Similar investigations found a connection between the proximal and final components of feedstock from biomass [17,15]. After that, Fig. 2 (b)–(d) showed the PCC values between inputs and outputs, which represent the corresponding significance of an input's characteristic in forecasting outputs.

With a PCC of 0.76, the input variable T has a significant impact. According to Fig. 2(b)'s negative PCC for T, increasing T would result in less biochar being produced. On the other hand, raising T might boost biochar's carbon content by bringing down volatile substances like H, O, and N, demonstrating the existence of a balance [15]. The effect of biochar's near compositions (FC, VM, and Ash) on input variables is shown in Fig. 2(c) The subsequent output vs. input combinations are shown below (in decreasing order) showed significant correlations: $|$ PCC$|$ $=$ 0.75 for VM vs. PT, 0.7 for Ash vs. ASH, 0.65 for FC vs. PT, and 0.4 for FC vs. ASH. Also, Fig. 2(d) demonstrated a significant impact of biochar extreme structure (C,H,O, N) on the accompanying (in diminishing request) input factors: $|$PCC$|$ rises to 0.8 for O versus PT, 0.75 for N versus N, 0.65 for H versus PT, 0.5 for C versus PT, and 0.5 for C versus Debris. Investigation of component importance found that PT altogether affected six of the eight result factors examined, including BYY, FCY, VMY, CY, HY, and OY. Both AshY and NY affected the other two factors, Debris and N. Accordingly, any vulnerabilities related with these factors would altogether impede the expectation capacity of the information driven models.

2.3. Proposed ML models

Multiple linear regression, a particular machine-learning model called Decision Tree, two ensemble methods called Adaboost Regressor and Bagging Regressor, as well as other factors were utilized to anticipate the biochar Yield. The development of ensemble models was built on the foundation of separate simulations. The same set of modeling parameters were used for both individual and ensemble models to create a baseline for contrast. Their wide range of applications, particularly in material engineering and building construction materials, served as the main justification for choosing these technologies[39,40]. The most recent Anaconda application uses Python programming as well as individual and collective ML approaches. All individual and ensemble models are run using the Python navigator's Spider (version 5.1.5). Such models are frequently used to forecast outcomes depending on the supplied input data.

2.3.1. Multiple linear regression

Regression models often calculate the strength of the correlation and describe the type of connection between the parameters of the input and output [41]. A common technique for determining the strength of the association between more than one variable is linear regression. MLR methods are types of models of regression that incorporate both independent and dependent variables [42]. A general linear connection is created as shown in Eq. (6) in order to determine the values of a factor that is dependent based on the values of the independent ones.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \bullet\bullet\bullet\bullet + \beta_n x_n \tag{6}$$

2.3.2. Decision tree

The DT is a categorization method incorporating categorization issues and complexities, with classes included inside the tree [43]. If there is no such thing as a class, the regression technique with independent parameters can be used to anticipate the outcome [44]. A hierarchical classifier, with inner nodes reflecting database properties, is what a DT is. While the branches show the decision rules, each leaf node displays the outcome. The two nodes that make up the DT are a leaf and a decision node. In contrast to leaf nodes, which only have a few number of branches, choice nodes can make any choice. The framework of this approach is shaped like a tree, with roots at the bottom and branches growing in every position [45]. The sets of data are categorized by the DT. The method evaluates the discrepancy between the intended and projected values at each division point. The gap between actual and expected results is calculated at each division point. The technique is repeated to locate the following split point. The location with the smallest difference between experimental and anticipated values is known as the split point. The diagram for the DT is shown in Fig. 3.

2.3.3. AdaBoost regressor

A artificial intelligence approach called the ensemble method is

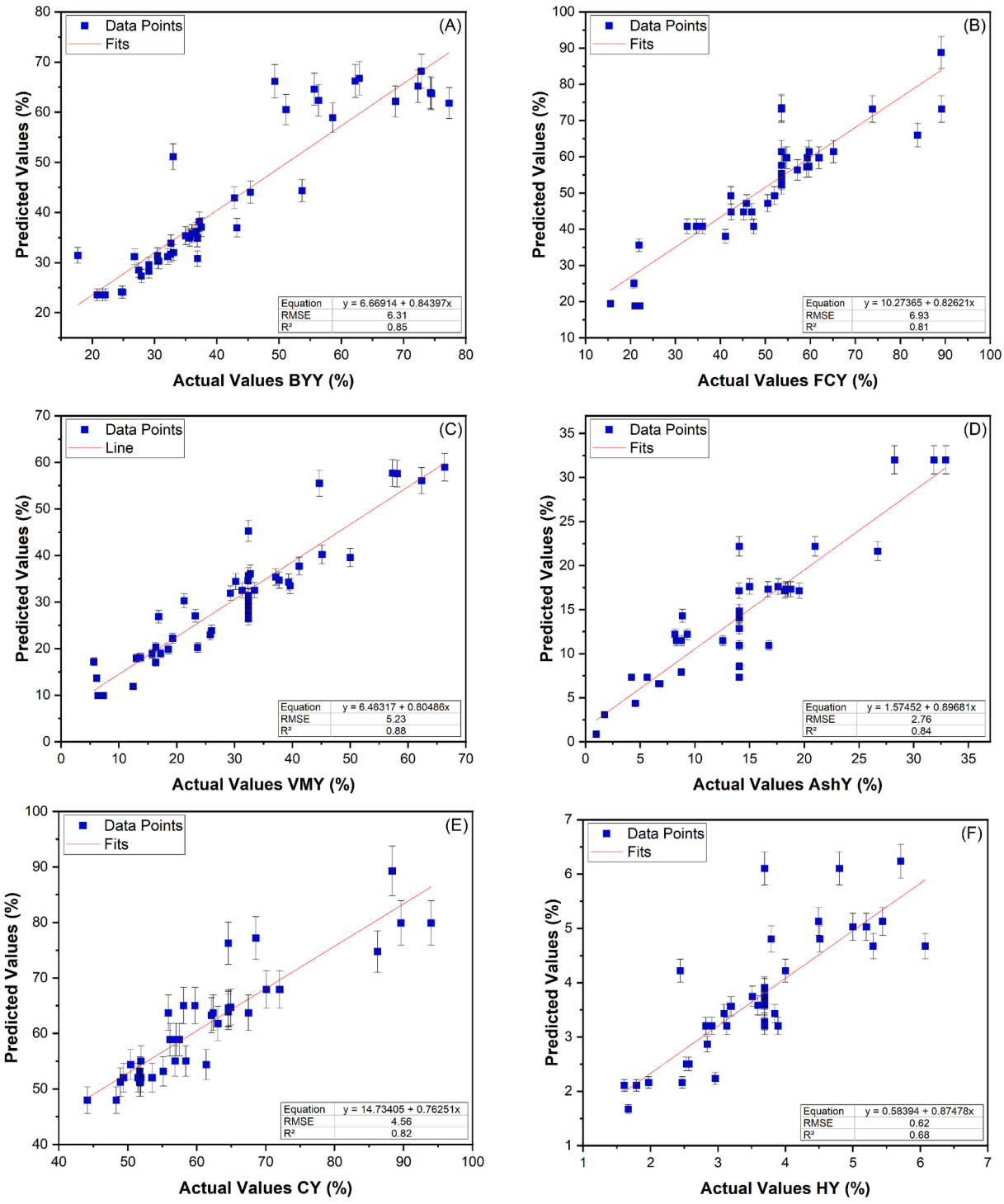

Fig. 7. Decision tree (DT) parity plots comparing actual and expected biochar yield, proximal composition (FC-VM-ash), and final composition (C—H—O—N).

utilized to optimize several models using just one method [46]. The entire set consists of numerous algorithms, commonly referred to as multi-classifiers. Many hundreds of thousands of students form an alliance and work together to solve the problem. Adaptive boosting is a different moniker for AR since each instance is given a new set of weights and examples that were mistakenly detected are given bigger weights. Boosting methods are extensively used in supervised ML to reduce variance and bias. The created data points serve as input data for another model. It utilized repeatedly until the necessary basic learners formed. The AR is most effective at improving DT performance on binary classification issues. It is also used to boost the performance of other ML algorithms. When applied to a slow learner, this technique performs admirably. Fig. 4 shows the whole method for anticipating the desired outcome of the AR algorithm.3.

2.3.4. *Bagging regressor*

When training the Bagging Regressor, extra data is added to the prediction model to improve accuracy. Certain observations can be reproduced in each by using sampling with replacement, a fresh training dataset is created. There is an equal chance that each stage of the

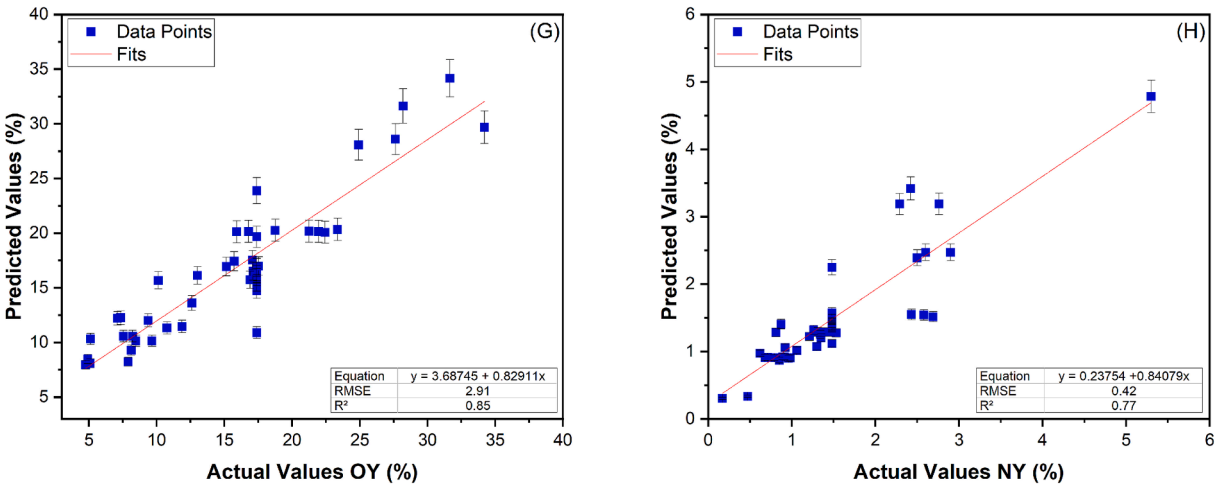

Fig. 7. (*continued*).

bagging procedure will appear in the freshly produced dataset. The capacity to predict is mostly unaffected by the size of the training batch of datapoints. In addition, by adjusting the expected outcome to the desired result, variation may be considerably reduced. The creation of new models frequently makes use of these data sets. The models in this ensemble average each forecast that is produced. Regression forecasts could be the average of the numbers of estimates from many models [47]. Twenty tiny models used to alter the DT with BR in order to identify the ideal configuration that produces a particular output result. The bagging method's process graph, shown in Fig. 5, shows the steps necessary to achieve the desired outcomes.

2.4. Model accuracy evaluation metrics

Two data-driven models' performance was examined and assessed on a dataset containing 226 data points using different training–testing splits. Consider two common measures for ML regression problems: R2 and RMSE, which were computed as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i^{exp} - Y_i^{pred}\right)^2} \tag{7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(Y_i^{exp} - Y_i^{pred}\right)^2}{\sum_{i=1}^{N}\left(Y_i^{exp} - Y_{ave}^{exp}\right)^2} \tag{8}$$

The estimated biochar yields from the experiment and the model are indicated as Y_i^{exp} and Y_i^{pred} respectively, while Y_{ave}^{exp} is the mean of all testing biochar outputs and N is the overall number of data points of data in this paper, which comes to 226.

2.5. Parameters analysis techniques

In this review, the SHAP (Shapley Added substance clarifications) worth and weight scores were utilized to uncover the weighting of different information boundaries on the projected qualities [48]. Higher weight scores for input boundaries show really displaying esteem. Weight scores, nonetheless, couldn't consider what the model's result meant for by the boundaries for this situation. A game-based hypothesis technique called SHAP was used to enhance the output of machine learning models. The SHAP esteem illustrates how a data boundary interacts with the model to derive the anticipated outcome from the foundation worth. Plotting the SHAP parameters of the information boundaries was necessary to visualize the coupling impact.

3. Results and Discussion

The study used four models namely MLR, DT, AR, and BR, were evaluated for their predictive performance. Among these models, the MLR model exhibited the lowest performance in terms of predictive accuracy When evaluated against the MLR, DT, AR, and BR models, the BR model performed the best. The evaluation of the models' performance based on two commonly used error analysis metrics i.e. R^2 and RMSE. Considering the presence of eight output parameters, the models assessed for their ability to predict these parameters accurately.

3.1. Predictive performance of multiple linear regression

The predictive accuracy of the MLR model was assessed using the R2 and RMSE measures, as illustrated in Fig. 6. For biochar yield prediction, the model's coefficient of R2 was 0.60, indicating that it might account for nearly 60 % of the variance in observed biochar yield. The model also performed well for the proximal composition parameters FCY, VMY, and AshY, with respective R2 coefficients of 0.37, 0.32, & 0.60. The equivalent RMSE figures were 11.27, 13.87, and 5.49. These findings indicate that the MLR model caught the connections between the input variables and the proximal composition characteristics adequately, with low average prediction errors.

The MLR model produced average R2 values of 0.41, 0.04, 0.33, and 0.33 for the final composition parameters CY, HY, OY, and NY, respectively. CY, HY, OY, and NY had average RMSE values of 7.88, 1.74, 6.24, and 6.24, respectively. Overall, the MLR model outperformed other models in capturing the complex relationships between input parameters and various output Parameters.

3.2. Predictive performance of DT

The parity graphs in Fig. 7 indicate the accuracy of the DT created in this work for forecasting. The evaluation of the DT model's predictive accuracy using R^2 and RMSE metrics showed that it achieved an R^2 and RMSE value of 0.85 and 6.31 for biochar yield (BYY), respectively. In predicting proximate composition, including FCY, VMY, and AshY, the DT model achieved R^2 values of 0.81, 0.88, and 0.84 with RMSE values of 6.93, 5.23 and 2.76, respectively. For the ultimate composition parameters including [43], HY, OY, and NY, the Decision Tree (DT) model exhibited moderate predictive performance, achieving average R^2 values of 0.82, 0.68, 0.85, and 0.77, along with corresponding RMSE values of 4.56, 0.62, 2.91, and 0.42, respectively. Despite these results, the DT model underperformed compared to the Adaboost Regressor (AR) and Bagging Regressor (BR) models, which demonstrated

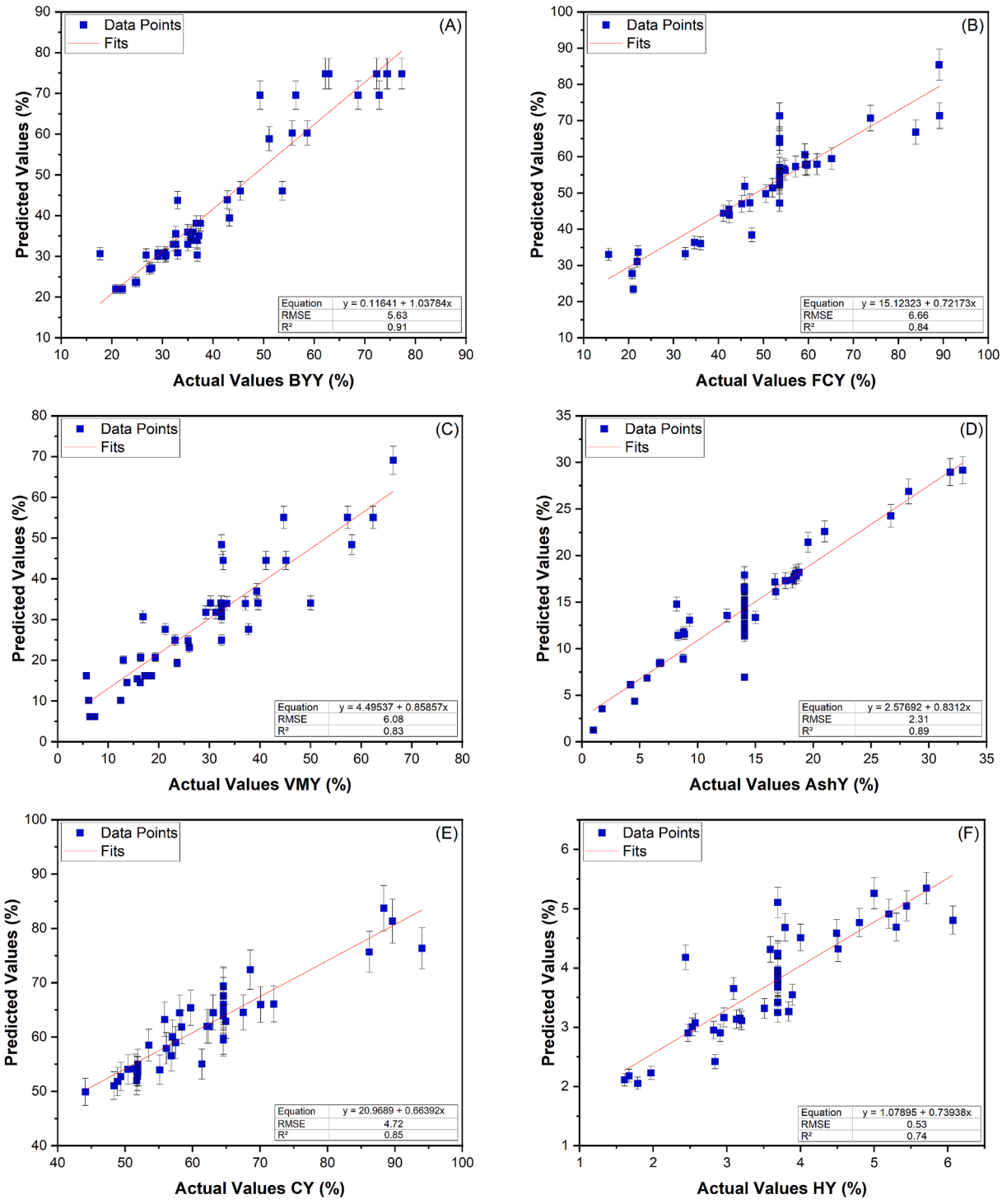

Fig. 8. Parity charts for Adaboost Regressor (AR) parity plots comparing the yield, proximal composition (FC-VM-ash), and ultimate composition (C—H—O—N) of biochar between actual and forecasted values are shown in the graphs below.

significantly higher predictive accuracy for all eight output parameters. The AR and BR models not only achieved higher R^2 values but also had lower RMSE scores, indicating their superior ability to capture the complex relationships between the input parameters and the multiple output variables. These results underscore the limitations of the DT model in accurately modeling intricate patterns in the dataset, making it less suitable for predicting biochar composition compared to the ensemble-based approaches.

3.3. Predictive performance of AR

The predictive accuracy of the AR model, as illustrated by the parity plots in Fig. 8, was evaluated using R^2 and RMSE metrics. In terms of biochar yield (BYY) prediction, the coefficient R2 of the AR model was 0.91, and the RMSE was 5.63. indicating its ability to account for approximately 91 % of the variance in the observed biochar yield. For the proximate composition parameters including FCY, VMY, and AshY, the AR model demonstrated consistent performance. It attained an R^2

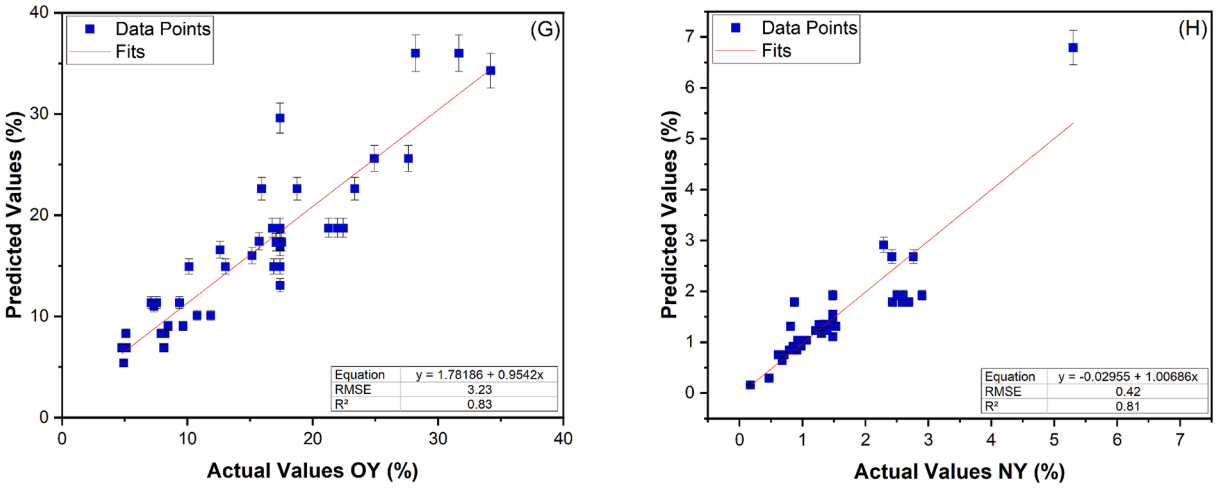

Fig. 8. (*continued*).

value of 0.84 for FC, 0.83 for VM, and 0.89 for ash content. The corresponding RMSE values measured at 6.66, 6.08, and 2.31, respectively.

These findings imply that the AR model caught the connections between the input variables and the proximal composition parameters adequately, with low average prediction errors. Regarding the ultimate composition parameters, (CY, HY, OY, and NY), the AR model exhibited favorable predictive accuracy. The average R2 values for the four parameters were 0.85, 0.74, 0.83, and 0.81 respectively. CY, HY, OY, and NY had average RMSE values of 4.72, 0.53, 3.23, and 0.42, respectively. In terms of prediction accuracy across all eight output metrics, the AR model beat the DT model overall. The AR model displayed greater performance in capturing the intricate relationships between the many output parameters and the input parameters, with consistently higher R2 values and lower RMSE values.

3.4. Predictive performance of BR

The BR model exhibited the highest predictive accuracy among the evaluated models, as evident from the parity plots visualized in Fig. 9. With an extraordinary R2 coefficient of 0.94 and a low RMSE coefficient of 4.27, the BR model successfully predicted biochar yield (BYY), showing that it could account for almost 94 % of the variance in observed biochar yield. In terms of the proximate composition parameters (FCY, VMY, and AshY), the BR model consistently displayed excellent performance. It achieved high R^2 values of 0.93 for FC, 0.93 for VM, and 0.94 for ash. The corresponding RMSE values were remarkably low at 4.35, 3.93, and 1.74, respectively. These results suggest that the BR model accurately captured the connections between the input parameters and the proximate composition parameters, exhibiting low average prediction errors. For the ultimate composition parameters (CY, HY, OY, and NY), the BR model demonstrated outstanding predictive accuracy. It achieved an R^2 value of 0.95, 0.90, 0.92 and 0.96 across all four parameters. The average RMSE values for CY, HY, OY, and NY recorded at 2.59, 0.32, 2.12, and 0.18 respectively, further highlighting the BR model's precision in predicting these parameters. Overall, the BR model outperformed both the AR and DT models in terms of predictive accuracy across all eight-output parameters. The BR model showed greater performance in capturing the complicated interactions between the input variables and the numerous output parameters, with frequently greater coefficients for R2 and lesser RMSE values. These results further emphasize the BR model's potential as a trustworthy model for such applications by demonstrating how well it can predict the relevant biochar properties.

3.5. Comparison and discussion

The accuracy of the various models developed for this study was assessed using statistical tests including RMSE and R2. This evaluation is summarised in Table 2. Table 2 displays the RMSE values for the training and testing datasets for DT, AR, and BR. The DT model's RMSE values were the greatest, whilst the BR model's were the lowest. This reveals that the BR model outperformed the AR and DT models for predicting important biochar features. This study set out to show how to estimate important biochar properties using both individual and ensemble machine learning techniques. To solve issues, the DT employs a tree model. The DT technique aims to build a model that can accurately predict a number of important factors. In supervised learning, boosting is used to lessen variation and bias. Its foundation is the idea that learning happens in stages. Every subsequent learner, with the exception of the first, is built from a predecessor. In some respects, having fewer makes your options stronger. On the contrary, bagging is a technique for choosing an arbitrary dataset variable using a training set using a different approach that enables the choice of new data values. The weakened models were taught separately after the generation of numerous data samples, and based on the task, the bulk or average among the values predicted produced a more precise forecast or estimation. We looked at the expected performance of each method employed in this study to see if the modelling technique was a superior predictor. The output of the BR model had the greatest R2 of 0.96 when compared to the other models, showing that it was more accurate [49–51]. To figure out the impact of each input parameter the anticipated values of the targeted biochar properties, parameter significance and effect analyses were also carried out. The performance of the model may be influenced by the parameters that are input and the number of data sets. Which of the four input parameters has the greatest influence on the anticipated outcome is shown by the parameter influence and significance.

3.6. Parameters analysis

3.6.1. Parameter importance

Fig. 10 shows parameter importance of input parameters and their weight ratings on each output parameter. This figure chosen to develop a thorough and accessible comprehension of the impact of different input parameters on prediction accuracy.

Results show that the temperature (T) had the most impact on output variables. Carbon content (C) highly influenced the CY, BYY, and AshY. Ash content (A) had a high impact on ASHY, FCY, CY, and VMY. Fig. 10, which also shows that T is the most sensitive variable, demonstrates the proportional contributions of Input variables (T, C, O, RT, HR, VM, Ash,

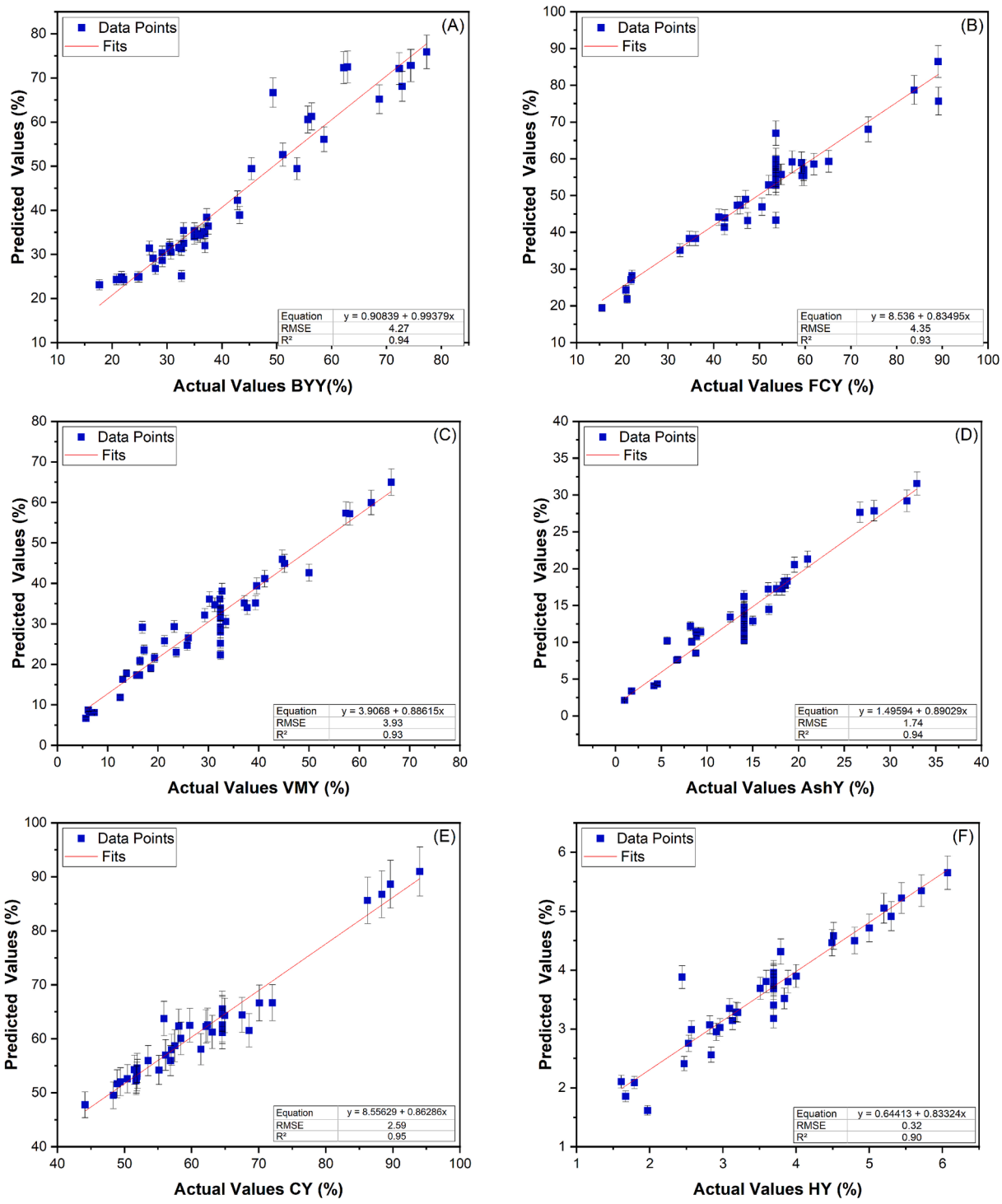

Fig. 9. Parity charts for Bagging Regressor (BR) parity plots comparing the yield, proximal composition (FC-VM-ash), and ultimate composition (C–H-O-N) of biochar between actual and forecasted values are shown in the graphs below.

N, H, and FC).

The results of this analysis show that the most critical input factors for determining the output variables are temperature and carbon content. Some output variables are also affected by ash content. These findings are utilised to enhance prediction model accuracy by ranking input characteristics based on their relevance.

3.6.2. *Parameters influence*

The SHAP values of important parameters were used during prediction to gain an increased awareness of how these variables affect model outcomes. The parameter's total impact on model output is represented by the mean of its SHAP values. This information may be used to get insights into the BR model by evaluating the SHAP values of the parameters judged to be most important. SHAP values, in particular, are used to investigate the effect of each parameter on a given prediction. A high SHAP value, for example, tends to increase the likelihood of a positive prediction, whereas a negative SHAP value tends to decrease the likelihood of a positive prediction. To assess a parameter's overall

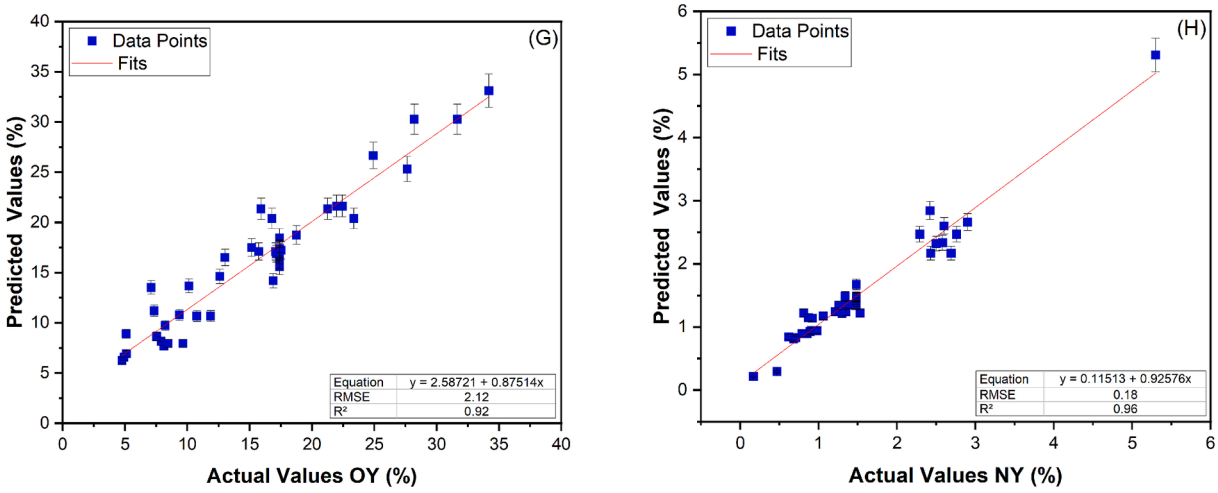

Fig. 9. (*continued*).

Table 2
R2 and RMSE values for biochar yield (BYY), proximate composition (FCY-VMY-AshY), and ultimate composition (CY-HY-OY-NY) prediction were obtained using varying quantities of training and testing data.

		R2				RMSE			
		DT	AR	BR	MLR	DT	AR	BR	MLR
BYY	Tra	0.88	0.88	0.94	0.64	5.22	5.07	3.47	9.31
	Tes	0.85	0.91	0.94	0.60	6.31	5.63	4.27	9.87
FCY	Tra	0.82	0.83	0.93	0.64	7.17	7.06	4.47	10.39
	Tes	0.81	0.84	0.93	0.37	6.93	6.66	4.35	11.27
VMY	Tra	0.84	0.86	0.93	0.70	6.84	6.33	4.43	9.83
	Tes	0.88	0.83	0.93	0.32	5.23	6.08	3.93	13.87
AshY	Tra	0.88	0.88	0.95	0.71	2.74	2.79	1.74	4.17
	Tes	0.84	0.89	0.94	0.60	2.76	2.31	1.74	5.49
CY	Tra	0.85	0.86	0.96	0.56	3.90	3.87	1.94	6.80
	Tes	0.82	0.85	0.95	0.41	4.56	4.72	2.59	7.88
HY	Tra	0.72	0.74	0.94	0.69	0.67	0.66	0.33	0.72
	Tes	0.68	0.74	0.90	0.04	0.62	0.53	0.32	1.74
OY	Tra	0.87	0.90	0.92	0.72	3.11	2.68	2.42	4.62
	Tes	0.85	0.83	0.92	0.33	2.91	3.23	2.12	6.24
NY	Tra	0.80	0.84	0.98	0.72	0.58	0.50	0.18	4.62
	Tes	0.77	0.81	0.96	0.33	0.42	0.42	0.18	6.24

influence on the model result, the parameter's SHAP value range average is employed.The SHAP values were employed in this study to get understanding of the BR model. The SHAP values of the most important parameters were examined, and it was determined that these parameters had a significant impact on model output. This information may be used to improve the BR model by identifying the parameters that have the most effect on the model output and then altering them accordingly. To understand the internal functioning of the GEP method, the SHAP parameters have been generated for all BR forecasts on the test split. The values for SHAP show how every attribute influenced the model's outcome forecast. A positive SHAP number indicates that the feature improved the prediction above the baseline value, whereas a negative SHAP number indicates that the feature degraded the predicted outcome.

Fig. 11 depicts the role of SHAP numbers. Colours ranging from dark red & light blue are employed to show the high and low values of features for every feature, respectively. The horizontal breakdown of the + ve & −ve SHAP numbers for each parameter is employed to understand the parameter's related SHAP value. The parameters are listed in descending order (from the top down to the bottom) depending on their respective relevance. In other words, variables with the highest actual SHAP numbers are crucial in the model's predictions. SHAP values are used to understand how the model predicts and to identify the features that are most essential to the model's performance.

Fig. 11 shows the key factors influencing the BR prediction. Light blue dots indicate lower T and Ash concentrations, which are associated with positive SHAP values in BYY, VMY, HY, and OY. Dark red dots indicate higher T and Ash concentrations, which are associated with highly positive SHAP values in AshY and CY. In FCY, higher T and lower ash values were associated with a highly positive SHAP value. High amounts of C in CY and N in NY were linked to positive SHAP numbers, whereas low levels elements were linked to negative SHAP values. NY's positive and negative SHAP levels were both moderately positive. This demonstrates that the influence of these small SHAP values on lowering and increasing prediction outcomes is much greater in BYY, VMY, HY, and OY, while it is the least in AshY, FCY, CY, and NY.

4. Conclusion

By presenting and testing a set of ensemble machine learning models for highly accurate biochar yield and composition prediction, this study makes a substantial contribution to the field of biochar manufacturing. The promise of ensemble techniques in improving prediction precision is demonstrated by the use of models like the Bagging Regressor, which performed better than more conventional models like Multiple Linear Regression (MLR), Decision Trees (DT), and Adaboost Regressor (AR). Specifically, the Bagging Regressor outperformed previous techniques, achieving an R2 value of up to 0.96 for biochar output. The model

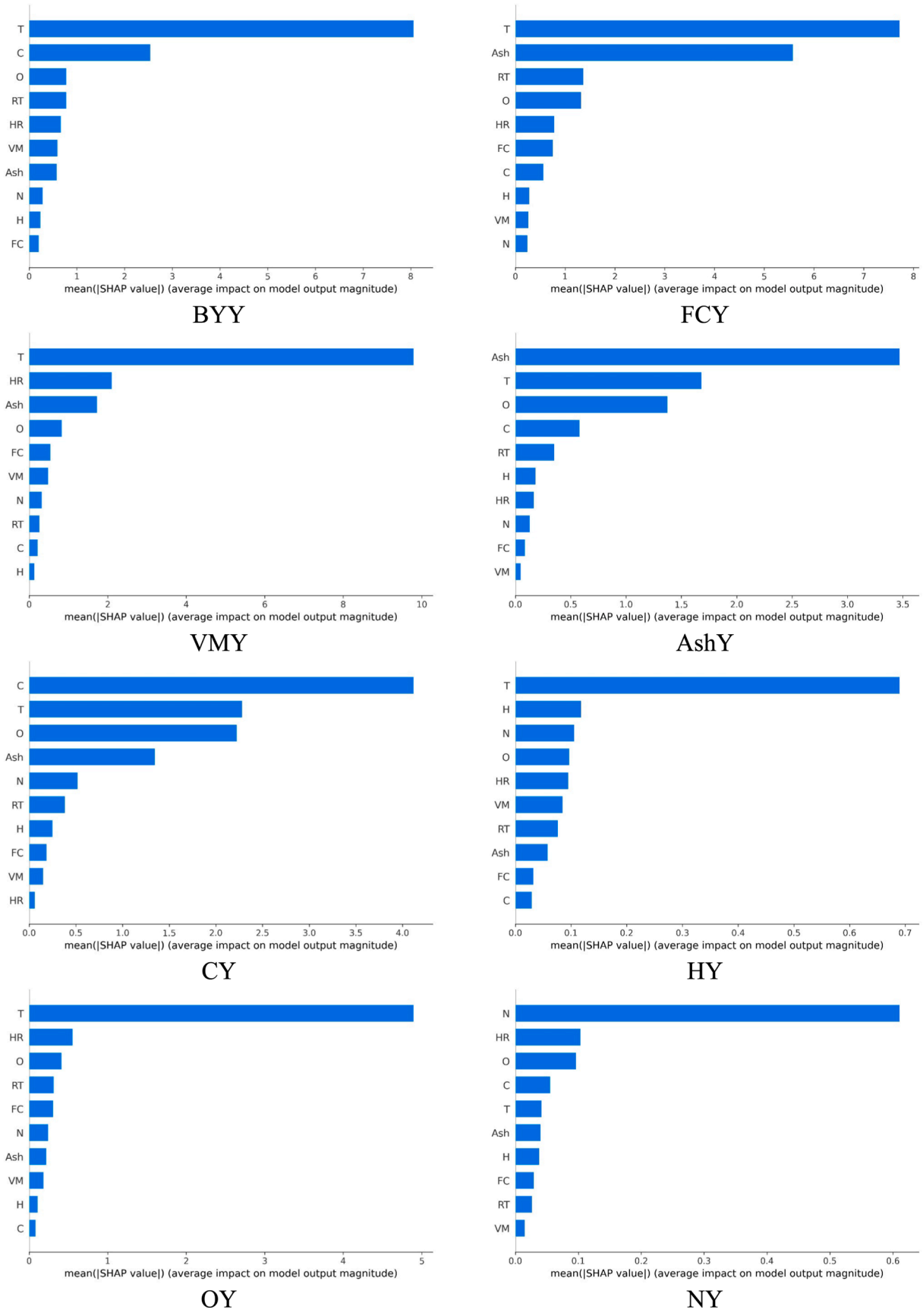

Fig. 10. Relative Contribution of output Parameters (BYY, FCY, VMY, AshY, CY, HY, OY, NY).

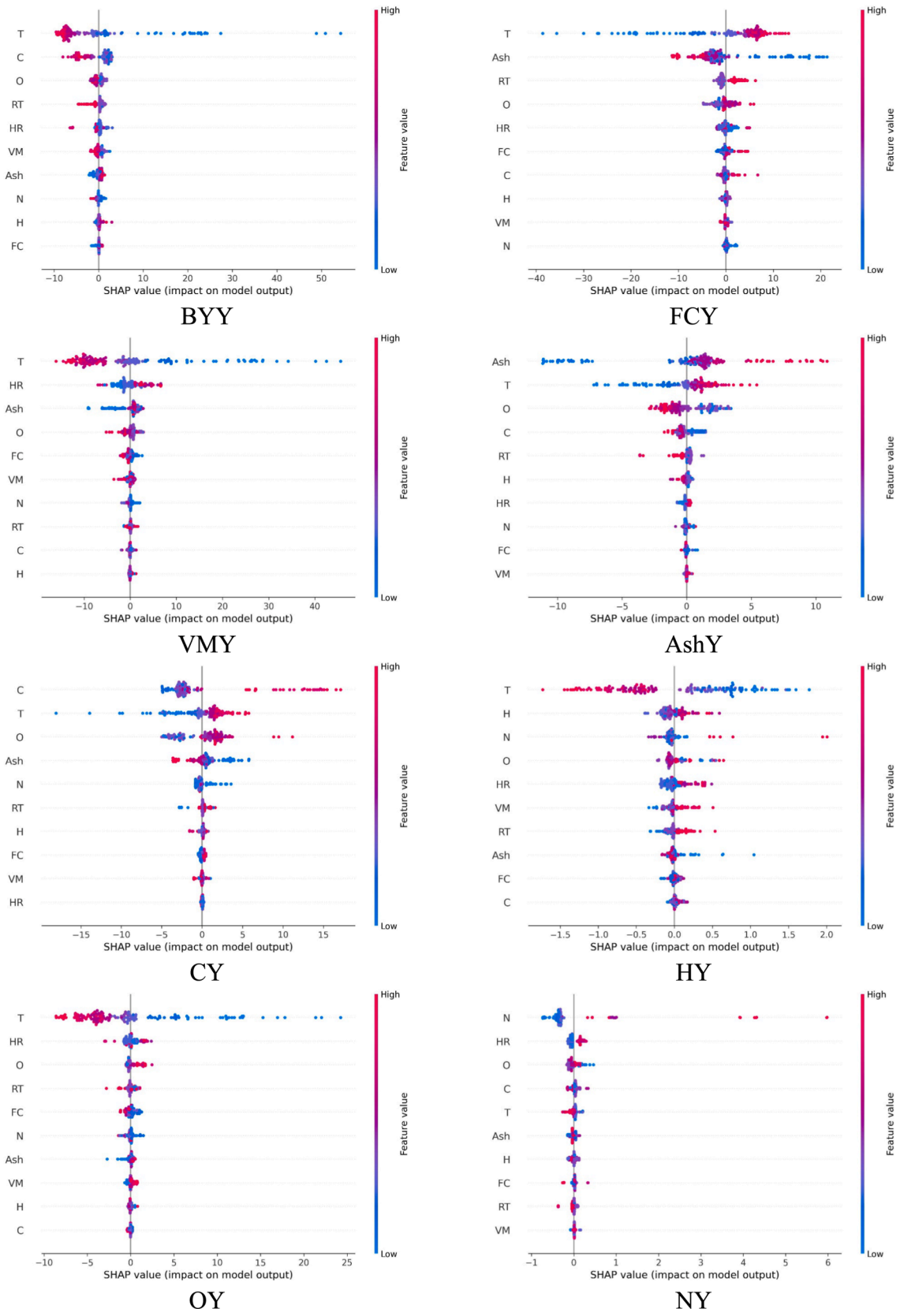

Fig. 11. SHAP parameter values demonstrating the effect of output parameters on the BR simulation (BYY, FCY, VMY, AshY, CY, HY, OY, NY).

provides more accurate yield and ultimate composition parameter predictions (such as Carbon Yield (CY), Hydrogen Yield, Oxygen Yield, and Nitrogen Yield (NY)) than traditional techniques, which adds scientific value. This research closes a significant gap in the literature on biochar modeling, as earlier attempts at modeling the material have been beset by poor computational efficiency and low generalizability.

The results of this study can potentially be used to optimize biochar production procedures in the industrial setting. The models presented here can be utilized by biochar producers to optimize their processes for various inputs, as they can accurately forecast biochar yield and composition across a range of biomass feedstocks and pyrolysis conditions. As a result, biochar may be used more effectively, more cheaply, and for better purposes like waste management, soil improvement, and carbon sequestration. The models can assist decision-making in a variety of biochar-related businesses thanks to their predictive capacity across many output characteristics, which encourages the use of more environmentally and agriculturally sustainable approaches.

Notwithstanding these efforts, the Study acknowledged several shortcomings. Despite being thorough with 226 datasets from published literature, the dataset employed in this study might not fully describe the heterogeneity of biochar synthesis processes, particularly when extreme pyrolysis conditions or uncommon feedstocks are involved. Additionally, methodological discrepancies in the data obtained prevented the inclusion of some biomass feedstocks and process characteristics, such as moisture content and particle size. Consequently, in order to address these unrepresented circumstances, the models might need to be further refined, which could affect their prediction ability in practical applications using novel or unusual feedstocks.

Given these constraints, there exist other potential directions for further investigation. Enhancing the dataset to encompass a wider variety of biomass kinds and pyrolysis settings would enhance the model's resilience and applicability. Moreover, utilizing cutting-edge machine learning methods like hybrid models or deep learning may improve prediction accuracy, especially for more intricate biochar compositions. Creating real-time predictive models that are integrated with industrial systems may also make it easier to dynamically optimize the processes involved in the manufacture of biochar, thus increasing efficiency. To further understand the effects of these and other crucial variables on biochar yield and composition, future research could include examine the effects of feedstock particle size, moisture content, and pyrolysis duration.

In conclusion, by presenting high-accuracy machine learning models that improve the predictability and efficiency of biochar production, this study makes a significant contribution to both academic research and industry application. The findings show how employing ensemble learning models in this field advances science and lay the groundwork for more research and development of biochar modeling methods. These models provide a viable tool for streamlining biochar manufacturing processes, encouraging sustainable farming methods, and assisting international environmental initiatives to manage organic waste and fight climate change by increasing the accuracy of yield and composition projections.

CRediT authorship contribution statement

Jingguo Gou: Conceptualization, Formal analysis, Funding acquisition. **Ghayas Haider Sajid:** Investigation, Resources, Software. **Mohanad Muayad Sabri:** Conceptualization, Formal analysis, Funding acquisition. **Mohammed El-Meligy:** Methodology, Resources, Supervision. **Khalil El Hindi:** Formal analysis, Investigation, Resources. **Nashwan Adnan OTHMAN:** Conceptualization, Methodology, Validation.

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.

Acknowledgement

The authors extend their appreciation to King Saud University for funding this work through Researchers Supporting Project number (RSPD2025R953), King Saud University, Riyadh, Saudi Arabia.

References

[1] Li Y, Ahmed A, Watson I, You S. Waste-to-biofuel and carbon footprints. Waste Biorefinery Integr. Biorefineries Waste Valoris 2020:579–97. https://doi.org/10.1016/B978-0-12-818228-4.00021-6.
[2] Tripathi N, Hills CD, Singh RS, Atkinson CJ. Biomass waste utilisation in low-carbon products: harnessing a major potential resource. Npj Clim Atmos Sci 2019;2. https://doi.org/10.1038/s41612-019-0093-5.
[3] Wijitkosum S, Sriburi T, Krutnoi L. Taking advantage of disposal bamboo chopsticks to produce biochar for greenhouse crop cultivation. Emerg Sci J 2024;8:917–32. https://doi.org/10.28991/ESJ-2024-08-03-07.
[4] Bolan N, Hoang SA, Beiyuan J, Gupta S, Hou D, Karakoti A, et al. Multifunctional applications of biochar beyond carbon storage. Int Mater Rev 2022;67:150–200. https://doi.org/10.1080/09506608.2021.1922047.
[5] Sajjadi B, Chen WY, Egiebor NO. A comprehensive review on physical activation of biochar for energy and environmental applications. Rev Chem Eng 2019;35:735–76. https://doi.org/10.1515/revce-2017-0113.
[6] Qiu B, Tao X, Wang H, Li W, Ding X, Chu H. Biochar as a low-cost adsorbent for aqueous heavy metal removal: a review. J Anal Appl Pyrolysis 2021;155. https://doi.org/10.1016/j.jaap.2021.105081.
[7] Dong JF, Wang QY, Guan ZW, Chai HK. High-temperature behaviour of basalt fibre reinforced concrete made with recycled aggregates from earthquake waste. J Build Eng 2022;48. https://doi.org/10.1016/j.jobe.2021.103895.
[8] Cha JS, Park SH, Jung SC, Ryu C, Jeon JK, Shin MC, et al. Production and utilization of biochar: a review. J Ind Eng Chem 2016;40:1–15. https://doi.org/10.1016/j.jiec.2016.06.002.
[9] Gwenzi W, Chaukura N, Wenga T, Mtisi M. Biochars as media for air pollution control systems: contaminant removal, applications and future research directions. Sci Total Environ. 2021;753. https://doi.org/10.1016/j.scitotenv.2020.142249.
[10] Barragán-Mantilla SP, Ramola S, Méndez A. Engineered biochar as a catalyst. Eng. Biochar Fundam. Prep. Charact. Appl. 2022:291–302. https://doi.org/10.1007/978-981-19-2488-0_15.
[11] Koestoer RH, Ligayanti T, Kartohardjono S, Susanto H. Down-streaming small-scale green ammonia to nitrogen-phosphorus fertilizer tablets for rural communities. Emerg Sci J 2024;8:625–43. https://doi.org/10.28991/ESJ-2024-08-02-016.
[12] Wang J, Ji J, Jiang Z, Sun L. Traffic flow prediction based on spatiotemporal potential energy fields. IEEE Trans Knowl Data Eng 2023;35:9073–87. https://doi.org/10.1109/TKDE.2022.3221183.
[13] Ma J, Zhang S, Liu X, Wang J. Machine learning prediction of biochar yield based on biomass characteristics. Bioresour Technol 2023;389. https://doi.org/10.1016/j.biortech.2023.129820.
[14] Li Y, Gupta R, You S. Machine learning assisted prediction of biochar yield and composition via pyrolysis of biomass. Bioresour Technol 2022;359. https://doi.org/10.1016/j.biortech.2022.127511.
[15] Leng L, Yang L, Lei X, Zhang W, Ai Z, Yang Z, et al. Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass. Biochar 2022;4. https://doi.org/10.1007/s42773-022-00183-w.
[16] Wang Z, Peng X, Xia A, Shah AA, Huang Y, Zhu X, et al. The role of machine learning to boost the bioenergy and biofuels conversion. Bioresour Technol 2022;343. https://doi.org/10.1016/j.biortech.2021.126099.
[17] Nguyen VG, Sharma P, Ağbulut Ü, Le HS, Tran VD, Cao DN. Precise prognostics of biochar yield from various biomass sources by Bayesian approach with supervised machine learning and ensemble methods. Int J Green Energy 2024;21:2180–204. https://doi.org/10.1080/15435075.2023.2297776.
[18] Xu Q, Du L, Deng R. Using machine learning to predict biochar yield and carbon content: enhancing efficiency and sustainability in biomass conversion. BioResources 2024;19:6545–58. https://doi.org/10.15376/biores.19.3.6545-6558.
[19] Liu Y, Dong J, Yuan S, Li K, Hu X, Wang Q. Variable fatigue loading effects on corrugated steel box girders with recycled concrete. J Constr Steel Res 2024;215. https://doi.org/10.1016/j.jcsr.2024.108526.
[20] Chen MW, Chang MS, Mao Y, Hu S, Kung CC. Machine learning in the evaluation and prediction models of biochar application: a review. Sci Prog 2023;106. https://doi.org/10.1177/00368504221148842.
[21] Mari Selvam S, Balasubramanian P. Influence of biomass composition and microwave pyrolysis conditions on biochar yield and its properties: a machine learning approach. Bioenergy Res 2023;16:138–50. https://doi.org/10.1007/s12155-022-10447-9.
[22] Bhattacharjee N, Biswas AB. Pyrolysis of orange bagasse: comparative study and parametric influence on the product yield and their characterization. J Environ Chem Eng 2019;7. https://doi.org/10.1016/j.jece.2019.102903.
[23] Biswas B, Pandey N, Bisht Y, Singh R, Kumar J, Bhaskar T. Pyrolysis of agricultural biomass residues: comparative study of corn cob, wheat straw, rice straw and rice

husk. Bioresour Technol 2017;237:57–63. https://doi.org/10.1016/j.biortech.2017.02.046.
[24] Chen D, Liu D, Zhang H, Chen Y, Li Q. Bamboo pyrolysis using TG-FTIR and a lab-scale reactor: analysis of pyrolysis behavior, product properties, and carbon and energy yields. Fuel 2015;148:79–86. https://doi.org/10.1016/j.fuel.2015.01.092.
[25] Chen D, Kang F, Li J, Zhu S, Liang X. Enhancement of underwater dam crack images using multi-feature fusion. Autom Constr 2024;167. https://doi.org/10.1016/j.autcon.2024.105727.
[26] Crombie K, Mašek O, Sohi SP, Brownsort P, Cross A. The effect of pyrolysis conditions on biochar stability as determined by three methods. GCB Bioenergy 2013;5:122–31. https://doi.org/10.1111/gcbb.12030.
[27] He X, Liu Z, Niu W, Yang L, Zhou T, Qin D, et al. Effects of pyrolysis temperature on the physicochemical properties of gas and biochar obtained from pyrolysis of crop residues. Energy 2018;143:746–56. https://doi.org/10.1016/j.energy.2017.11.062.
[28] Hong Z, Zhong F, Niu W, Zhang K, Su J, Liu J, et al. Effects of temperature and particle size on the compositions, energy conversions and structural characteristics of pyrolysis products from different crop residues. Energy 2020;190. https://doi.org/10.1016/j.energy.2019.116413.
[29] Lee Y, Park J, Ryu C, Gang KS, Yang W, Park YK, et al. Comparison of biochar properties from biomass residues produced by slow pyrolysis at 500°C. Bioresour Technol 2013;148:196–201. https://doi.org/10.1016/j.biortech.2013.08.135.
[30] Liu X, Zhang Y, Li Z, Feng R, Zhang Y. Characterization of corncob-derived biochar and pyrolysis kinetics in comparison with corn stalk and sawdust. Bioresour Technol 2014;170:76–82. https://doi.org/10.1016/j.biortech.2014.07.077.
[31] Liu Z, Han G. Production of solid fuel biochar from waste biomass by low temperature pyrolysis. Fuel 2015;158:159–65. https://doi.org/10.1016/j.fuel.2015.05.032.
[32] Liu Z, Niu W, Chu H, Niu Z. Process optimization for straws pyrolysis and analysis of biochar physiochemical properties. Nongye Gongcheng Xuebao/Transactions Chinese Soc Agric Eng 2018;34:196–203. https://doi.org/10.11975/j.issn.1002-6819.2018.05.026.
[33] Liu Z, Niu W, Chu H, Zhou T, Niu Z. Effect of the carbonization temperature on the properties of biochar produced from the pyrolysis of crop residues. BioResources 2018;13:3429–46. https://doi.org/10.15376/biores.13.2.3429-3446.
[34] Liu Z, Zhang F, Liu H, Ba F, Yan S, Hu J. Pyrolysis/gasification of pine sawdust biomass briquettes under carbon dioxide atmosphere: study on carbon dioxide reduction (utilization) and biochar briquettes physicochemical properties. Bioresour Technol 2018;249:983–91. https://doi.org/10.1016/j.biortech.2017.11.012.
[35] Patra BR, Nanda S, Dalai AK, Meda V. Slow pyrolysis of agro-food wastes and physicochemical characterization of biofuel products. Chemosphere 2021;285. https://doi.org/10.1016/j.chemosphere.2021.131431.
[36] Rout T, Pradhan D, Singh RK, Kumari N. Exhaustive study of products obtained from coconut shell pyrolysis. J Environ Chem Eng 2016;4:3696–705. https://doi.org/10.1016/j.jece.2016.02.024.
[37] Shariff A, Noor N, Lau A, Ali MAM. A comparative study on biochar from slow pyrolysis of Corn Cob and Cassava wastes. Int J Biotechnol Bioeng 2016;10:767–71.
[38] Tag AT, Duman G, Ucar S, Yanik J. Effects of feedstock type and pyrolysis temperature on potential applications of biochar. J Anal Appl Pyrolysis 2016;120:200–6. https://doi.org/10.1016/j.jaap.2016.05.006.
[39] Ucar S, Ozkan AR. Characterization of products from the pyrolysis of rapeseed oil cake. Bioresour Technol 2008;99:8771–6. https://doi.org/10.1016/j.biortech.2008.04.040.
[40] Zhang Y, Gao Z, Wang X, Liu Q. Predicting the pore-pressure and temperature of fire-loaded concrete by a hybrid neural network. Int J Comput Methods 2022;19. https://doi.org/10.1142/S0219876221420111.
[41] Khan M, Ullah Z, Mašek O, Raza Naqvi S, Nouman Aslam Khan M. Artificial neural networks for the prediction of biochar yield: a comparative study of metaheuristic algorithms. Bioresour Technol 2022;355. https://doi.org/10.1016/j.biortech.2022.127215.
[42] Song Y, Huang Z, Jin M, Liu Z, Wang X, Hou C, et al. Machine learning prediction of biochar physicochemical properties based on biomass characteristics and pyrolysis conditions. J Anal Appl Pyrolysis 2024;181. https://doi.org/10.1016/j.jaap.2024.106596.
[43] Ali A, Iqbal MJ. National disaster management act of Pakistan: a review. J Disaster Emerg Res 2010;2021. https://doi.org/10.18502/jder.5810.
[44] Kang F, Wu Y, Ma J, Li J. Structural identification of super high arch dams using Gaussian process regression with improved salp swarm algorithm. Eng Struct 2023;286. https://doi.org/10.1016/j.engstruct.2023.116150.
[45] Chen D, Kang F, Chen J, Zhu S, Li H. Effect of light source wavelength on surface defect imaging in deep-water concrete dams. NDT E Int 2024;147. https://doi.org/10.1016/j.ndteint.2024.103198.
[46] Farooq F, Ahmed W, Akbar A, Aslam F, Alyousef R. Predictive modeling for sustainable high-performance concrete from industrial wastes: a comparison and optimization of models using ensemble learners. J Clean Prod 2021;292. https://doi.org/10.1016/j.jclepro.2021.126032.
[47] Alkhattabi L, Arif K. Novel base predictive model of resilient modulus of compacted subgrade soils by using interpretable approaches with graphical user interface. Mater Today Commun 2024;40. https://doi.org/10.1016/J.MTCOMM.2024.109764.
[48] Javed MF, Farooq F, Memon SA, Akbar A, Khan MA, Aslam F, et al. New prediction model for the ultimate axial capacity of concrete-filled steel tubes: an evolutionary approach. Crystals 2020;10:741. https://doi.org/10.3390/cryst10090741.
[49] Alyousef R, Nassar RUD, Khan M, Arif K, Fawad M, Hassan AM, et al. Forecasting the strength characteristics of concrete incorporating waste foundry sand using advance machine algorithms including deep learning. Case Stud Constr Mater 2023;19. https://doi.org/10.1016/j.cscm.2023.e02459.
[50] Dong JF, Xu Y, Guan ZW, Wang QY. Freeze-thaw behaviour of basalt fibre reinforced recycled aggregate concrete filled CFRP tube specimens. Eng Struct 2022;273. https://doi.org/10.1016/j.engstruct.2022.115088.
[51] Huang J, Feng C, Wang X, Zhang Y. Continuous–discontinuous element method for simulating three-dimensional reinforced concrete structures. Struct Concr 2024. https://doi.org/10.1002/suco.202300531.