Bioresource Technology 359 (2022) 127511

Contents lists available at ScienceDirect

Bioresource Technology

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

Machine learning assisted prediction of biochar yield and composition via pyrolysis of biomass

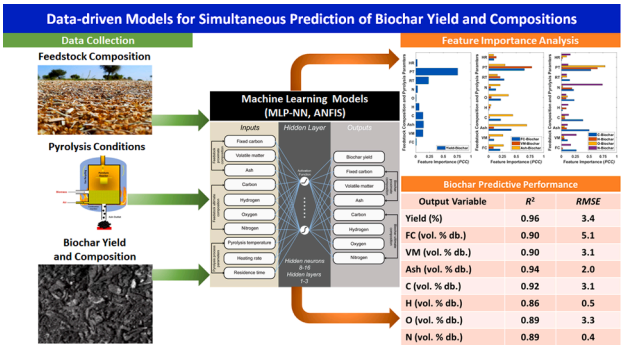
Yize Li, Rohit Gupta, Siming You *

James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK

HIGHLIGHTS

- Machine learning predict biochar production from pyrolysis of organic waste.
- Biochar yield, proximate and ultimate compositions are predicted simultaneously.
- Statistical analysis of various feedstock and biochar properties are performed.
- Neural network outperforms fuzzy inference system for prediction purposes.
- Sensitivity analysis reveals influence of optimal model parameters.

GRAPHICAL ABSTRACT



Output Variable	R^2	RMSE
Yield (%)	0.96	3.4
FC (vol. % db.)	0.90	5.1
VM (vol. % db.)	0.90	3.1
Ash (vol. % db.)	0.94	2.0
C (vol. % db.)	0.92	3.1
H (vol. % db.)	0.86	0.5
O (vol. % db.)	0.89	3.3
N (vol. % db.)	0.89	0.4

ARTICLE INFO

Keywords:
Neural Networks
Fuzzy Inference System
Organic Waste
Bioenergy
Sustainable Development

ABSTRACT

Biochar production via pyrolysis of various organic waste has potential to reduce dependence on conventional energy sources and mitigate global warming potential. Existing models for predicting biochar yield and compositions are computationally-demanding, complex, and have low accuracy for extrapolative scenarios. Here, two data-driven machine learning models based on Multi-Layer Perceptron Neural Network and Artificial Neuro-Fuzzy Inference System are developed. The data-driven models predict biochar yield and compositions for a variety of input feedstock compositions and pyrolysis process conditions. Feature importance assessment of the input dataset revealed their competitive significance for predicting biochar yield and compositions. Overall, the predictive accuracy of the models was up to 12.7% better than the Random Forest and eXtreme Gradient Boosting machine learning algorithms reported in the literature. The models developed are valuable for environmental footprint assessment of biochar production and rapid system optimization.

1. Introduction

With the growing concern about the global energy crisis and climate change, finding alternative energy sources has become increasingly important, and biomass is emerging as a promising sustainable energy source (Li et al., 2020). Biomass is produced at around 140 billion metric tonnes per annum globally, and its improper disposal causes environmental pollution (Tripathi et al., 2019). Organic waste, a major type of

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

https://doi.org/10.1016/j.biortech.2022.127511
Received 17 May 2022; Received in revised form 20 June 2022; Accepted 21 June 2022
Available online 22 June 2022

waste biomass, has a high potential for producing biochar via thermochemical processes. Biochar is gaining interest as one of the most valuable renewable bioresources, having a wide range of applications such as pollutant adsorption, greenhouse gas (GHG) reduction, wastewater treatment, soil remediation, energy production, and usage as catalysts (Bolan et al., 2021).

Biochar has distinctive properties such as large specific surface area, high mineral content, and surface functional groups. Specifically, it is used in a variety of applications due to its carbon-rich microporous structure (Sajjadi et al., 2019). Biochar has been applied as an adsorbent for the adsorption of water and air pollutants, where the yield is dictated by the type of feedstocks and choice of production process e.g., pyrolysis, gasification etc. (Qiu et al., 2021). For water pollution adsorption, biochar has been applied to remove heavy metals in wastewater. Literature suggests that biochar produced from rice husk and manure as feedstocks effectively removed Pb, Cu, Zn and Cd from aqueous solutions (Cha et al., 2016). It was also found that biochar is effective in the removal of metal vapours, especially elemental mercury (Hg^0), acid gases, ozone, nitrogen oxides and organic pollutants including aromatic compounds, volatile organic compounds, and odorous substances (Gwenzi et al., 2021). Another essential property is the catalytic nature of biochar which facilitates various applications such as syngas purification, biodiesel production, and air pollution control (Lee et al., 2017). Besides, biochar offers exceptional benefits for soil remediation. Part of the carbon in biochar is recalcitrant and could persist in the soil environment for an extended period, resulting in the sequestration of carbon upon soil application (Duan et al., 2020).

Pyrolysis is one of the widely deployed technologies for biochar production. The process involves heating organic matter in an inert atmosphere (i.e., in the absence of oxygen) to produce biochar, bio-oil, and syngas as value-added products (Irfan et al., 2016). The impacts of biochar application on soil quality and carbon abatement are dictated by the pyrolysis process parameters, and types and composition of biomass feedstocks. For example, higher temperatures, higher heating rates, lower pressures, or smaller particle sizes promote the decomposition of biomass polymer and reduce the yield of biochar from pyrolysis (Lee et al., 2013). On the other hand, higher pyrolysis temperature increases the carbon content and thus the quality of biochar for soil application. This trade-off indicates that there exist optimal pyrolysis process conditions towards biochar production (Brassard et al., 2018). Hence, the capability of understanding and predicting the yield is critical for designing optimal biochar systems for sustainable agricultural and chemical applications.

Over the past decades, theoretical models have been widely used for predicting the yields of pyrolysis processes (Kaczor et al., 2020). However, these models are often complex, time-consuming, and of semi-empirical nature (depending on the use of experimental data), which limits their use for biochar yield prediction towards more sophisticated process design and optimization (e.g., multi-objective optimization of biochar implementation). Researchers have also developed empirical correlations based on the predictions from the theoretical models or experiments (Wang et al., 2022). However, these empirical correlations are constricted to a limited range of experimental conditions and biomass feedstocks, and they are normally not suitable for extrapolative scenarios.

With the advent of artificial intelligence and an abundance of pyrolysis experimental data, data-driven modelling has become a popular method for predicting biochar production. These methods have superior prediction accuracy, shorter computation time, and complex data trend-reproducing capabilities (Wang et al., 2022). In particular, the method could effectively predict biochar production through limited experimental and system data. It finds the relationship between input and output variables through training and produces results without any *a priori* assumptions. However, the accuracy of existing Machine Learning (ML)-based biochar prediction models is limited especially when a small pool of datasets was used for the model development. For example, prior work has developed a Random Forest (RF) regression-based biochar yield and carbon content prediction model using 245 datasets that covered various biomass feedstocks and process operating conditions. The work achieved a coefficient of determination (R^2) value of 0.855 and 0.848 for predicting biochar yield and carbon content, respectively (Zhu et al., 2019). Another work explored the accuracy of the eXtreme Gradient Boosting (XGB) ML model for predicting biochar yield, which had an R^2 value of 0.84 based on 91 training datasets (Pathy et al., 2020). A recent work achieved an improved $R^2 = 0.92$ for predicting biochar yield using an artificial neural network (ANN) coupled with metaheuristic algorithms (Khan et al., 2022). However, there have been no prior efforts for developing a comprehensive data-driven model that can simultaneously predict biochar yield and compositions (both proximate and ultimate). As evidenced in the literature, the yield and compositions of biochar can have significant trade-offs based on the choice of pyrolysis conditions and biomass feedstock compositions. Therefore, it is of significant scientific interest to develop a comprehensive ML model that can predict both biochar yield and composition, which serves as the purpose of the present research work.

Multi-Layer Perceptron Neural Network (MLP-NN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) are two popular prediction methods. MLP-NN is a widely used data-driven modelling method in pattern recognition, signal processing, function approximation, and process simulation. They are often used to model complex relationships between inputs and outputs parameter spaces, explore patterns in data, or capture statistical structure in unknown joint probability distributions between observed variables (Asgari et al., 2021c). In contrast, ANFIS is a combination of adaptive control techniques, ANN, and fuzzy inference systems. As a deep learning algorithm, fuzzy logic allows the ambiguity of human perception or decision-making to be represented as a mathematical model (Abd El Aziz et al., 2017). The application of the models toward biochar production prediction is still limited.

This study envisions the significance of developing data-driven models to predict biochar yield, proximate composition, and ultimate composition. To the best of the author's knowledge, this is the first work that develops a comprehensive model to predict biochar yield and composition, simultaneously, given the pyrolysis conditions and biomass feedstock compositions. The work also highlights a comparison of the predictive performances of two different data-driven models: MLP-NN and ANFIS. During the data assimilation stage, feedstock compositions of various types of organic waste (mainly straw- and wood-based), related pyrolysis process parameters, biochar yields, and biochar compositions have been considered. The influences of different model parameters and training–testing dataset split are investigated through two metrics, i.e. the root mean squared error (RMSE) and R^2. Finally, the findings are discussed and concluded, while indicating areas for future improvements.

2. Materials and methods

2.1. Data collection and preprocessing

In total, 226 datasets were collected from the literature (19 studies) to develop the data-driven models, summarised in Table 1. The dataset includes a wide range of feedstocks such as Corncob, Corn stover, Bagasse, Cocopeat, Coconut shell, Coconut fibre, Wheat straw, Rice husk, Rice straw, Pine, Pine sawdust, Pine wood, Bamboo, Orange bagasse, Orange pomace, Rape stalk, Cassava stem, Cassava rhizome, Cotton stalk, Palm kernel shell, Wood stem, Wood bark, Agro-food waste, Canola hull, Oat hull, Vine pruning, Poultry litter, and Hinoki cypress (Bhattacharjee & Biswas, 2019; Biswas et al., 2017; Chen et al., 2015; Crombie & Mašek, 2015; Crombie et al., 2013; He et al., 2018; Hong et al., 2020; Lee et al., 2013; Liu et al., 2014; Liu & Han, 2015; Liu et al., 2018a; Liu et al., 2018b; Liu et al., 2018c; Patra et al., 2021; Rout et al., 2016; Shariff et al., 2016; Tag et al., 2016; Ucar & Ozkan, 2008; Zhang et al., 2017). To ensure generalizability, various attributes were

Table 1
Statistical summary of input and output variables for the raw dataset.

I/O	Type of variable	Variable	Min	Max	Mean	SD	No. of samples (% of total dataset)
Input	Feedstock proximate composition	FC (% db.)	4.33	27.80	13.84	5.36	226 (100%)
		VM (% db.)	68.20	91.16	79.83	4.91	226 (100%)
		ash (% db.)	0.16	15.14	6.33	3.94	226 (100%)
	Feedstock ultimate composition	C (% db.)	35.70	64.23	44.19	5.44	226 (100%)
		H (% db.)	4.10	10.18	5.94	0.99	226 (100%)
		O (% db.)	27.61	53.10	42.44	5.20	210 (92.9%)
		N (% db.)	0.00	9.61	1.29	1.66	226 (100%)
		S (% db.)	0.09	0.92	0.48	0.24	152 (67.3%)
	Feedstock lignocellulosic composition	Cel (% db.)	17.89	47.67	37.52	8.24	122 (53.9%)
		Hem (% db.)	11.48	56.29	24.97	13.30	117 (51.7%)
		Lig (% db.)	4.99	32.26	22.56	6.54	122 (53.9%)
	Pyrolysis condition	RT (min)	1.00	90.00	38.08	18.76	226 (100%)
		PT (°C)	200	800	460.7	124.8	226 (100%)
		HR (°C/min)	5.00	25.00	11.37	5.69	226 (100%)
Output	Process Efficiency	Biochar yield (%)	17.68	95.89	39.53	15.10	226 (100%)
		Energy yield (%)	38.40	99.80	60.07	14.66	87 (38.5%)
		HHV (MJ/kg)	3.6	37.66	23.28	5.50	105 (46.5%)
	Biochar proximate composition	FC (% db.)	15.04	94.11	53.57	19.77	159 (70.4%)
		VM (% db.)	0.49	82.72	32.37	20.00	159 (70.4%)
		ash (% db.)	0.32	37.91	14.05	9.02	159 (70.4%)
	Biochar ultimate composition	C (% db.)	44.12	94.61	64.57	12.07	162 (71.7%)
		H (% db.)	1.26	8.72	3.69	1.50	150 (66.3%)
		O (% db.)	0.00	45.17	17.39	10.27	150 (66.3%)
		N (% db.)	0.00	9.05	1.48	1.38	162 (71.7%)
		S (% db.)	0.00	1.29	0.53	0.31	115 (50.9%)

considered during the data collection stage, which included (1) proximate composition of biomass feedstock (2) ultimate composition of biomass feedstock, (3) lignocellulosic composition of biomass feedstock, (4) major pyrolysis conditions, (5) biochar yield, (6) proximate composition of biochar, (7) ultimate composition of biochar, (8) higher heating value (HHV) of biochar, and (9) energy yield of biochar. Although particle size of feedstock has been considered in prior works (Khan et al., 2022; Zhu et al., 2019), it was not included in the present work due to significant methodological inconsistencies and uncertainties associated with the data collection phase.

The proximate compositions for both the biomass feedstocks and biochar were expressed in dry basis having the following components: fixed carbon (FC), volatile matter (VM), and ash. It is important to note that the scope of the present dataset is limited to the feedstock ash content range 0–15%. To develop data-driven models for higher ash content input data, further expansion of the dataset is required in the future. The ultimate (or elemental) composition for the feedstocks and biochar is expressed in terms of carbon, hydrogen, nitrogen, oxygen, and sulphur (C-H-O-N-S). It is important to note that the literature contained a mix of wet-basis and dry-basis data for both feedstocks and biochar, which was converted on a dry-basis in the work using Eq. (1), where MC denotes the moisture content. The lignocellulosic composition of the feedstock contained three components such as lignin (Lig), cellulose (Cel), and hemicellulose (Hem). Essential pyrolysis process parameters found in the literature were pyrolysis temperature (PT), heating rate (HR), and residence time (RT). The complete dataset is available in the Supplementary Material.

$$\begin{aligned} \mathrm{FC_{dry}} &= \frac{\mathrm{FC_{wet}}}{1-\mathrm{MC}} \\ \mathrm{VM_{dry}} &= \frac{\mathrm{VM_{wet}}}{1-\mathrm{MC}} \\ \mathrm{ash_{dry}} &= \frac{\mathrm{ash_{wet}}}{1-\mathrm{MC}} \end{aligned} \tag{1}$$

Since the data collection was performed based on a wide range of literature, there were inevitably inconsistencies in the datasets, leading to missing values (see Table 1). For the input parametric space, several attributes were dropped if less than 70% of the data was available. Based on this, the sulphur content and lignocellulosic composition of feedstock were excluded from model development. For the output dataset, the cut-off criteria were set to 65%, leading to exclusion of sulphur content in biochar, HHV, and energy yield of biochar. Consequently, the modified dataset used for developing the model contained 10 input variables related to biomass feedstock compositions and pyrolysis conditions ($\mathrm{FC_{FS}}, \mathrm{VM_{FS}}, \mathrm{ash_{FS}}, \mathrm{C_{FS}}, \mathrm{H_{FS}}, \mathrm{O_{FS}}, \mathrm{N_{FS}}$, PT, HR, and RT), and 8 output variables quantifying biochar yield and compositions ($\mathrm{Yield_{BC}}, \mathrm{FC_{BC}}, \mathrm{VM_{BC}}, \mathrm{ash_{BC}}, \mathrm{C_{BC}}, \mathrm{H_{BC}}, \mathrm{O_{BC}}$, and $\mathrm{N_{BC}}$). The modified dataset is available in the Supplementary Material.

The modified dataset still contained several missing values, which may lead to erroneous model training. To circumvent this problem, the missing values corresponding to an attribute were substituted by the mean of the attribute (Ascher et al., 2022), which further ensured a continuous dataset. This process was only done during model training and therefore would not affect the performance of model testing.

Since the dataset contained variables of different ranges (i.e., difference between maximum and minimum values), mean, and standard deviation (SD) as shown in Table 1, data normalization was performed as an essential pre-processing step. Following conventional practices in ML, the standard normal variate Z_i was used, expressed as follows.

$$Z_i = \frac{X_i - \rho}{\sigma} \tag{2}$$

where X_i is the raw data, ρ is the mean, and σ is the SD of a variable.

To assess the relationship among any two variables (either between two different inputs or between inputs and outputs) Pearson Correlation Coefficient (*PCC*) was used, which quantify the degree of linear dependence (see Eq. (3)) (Zhu et al., 2019). A $PCC = 1$ or $PCC = -1$ suggests that the variables are highly correlated, while a $PCC = 0$ means that the two variables are not correlated. This absolute value of *PCC* also revealed the relative importance of feature that impacts the output variables such as biochar yield, proximate composition (FC-VM-ash), and ultimate composition (C-H-O-N). Results obtained from the *PCC* and relative importance of various features are discussed in Section 3.1.

$$PCC = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\sum_{i=1}^{n}(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{3}$$

Here, $\bar{x}$ and $\bar{y}$ are the two variables of interest among which *PCC* is to be

determined and n is the number of datapoints.

2.2. Artificial Neuro-Fuzzy Inference System

ANFIS is based on the Takagi-Sugeno inference method, which creates a nonlinear mapping from the input space to the output space using IF-THEN rules. The IF-THEN rules represent the form of "if A and B, then C" as presented in Table 2. The ANFIS model combines the characteristics of ANN and the fuzzy inference system (Bakyani et al., 2016).

The ANFIS model has five layers. The interactions between the different layers of the ANFIS model are shown in Fig. 1. The first layer nodes represent the input variables (i.e., proximate composition of feedstock, ultimate composition of feedstock, HR, PT, and RT) and fuzzification. Each node in this layer is a fuzzy set and any output from it is the membership degree given from the membership function (MF) of the fuzzy set. The node of layer 2 is the rule node. This layer calculates the degree of MF by activating the AND operator of IF-THEN rules. Each output node in this layer represents the trigger strength of each rule. The third layer is responsible for determining the relative value of each rule in relation to the firing strength. The result of this layer is called the normalised firing strength. Nodes of layer 4 uses the consequent part of the Takagi-Sugeno approach, where subsequent parameter values and normalised firing strengths from the third layer of the rule base are calculated. The layer 5 is the output node, which corresponds to biochar yield, proximate composition, and ultimate composition. A simplified mathematical formulation of ANFIS is shown in Table 2. Here $A1$, $A2$ and $B1$, $B2$ are MFs of inputs X and Y, which are used to manipulate variables (or known as fuzzification). p_1, q_1, r_1 and p_2, q_2, r_2 are the relevant parameters of the output function determined during the model training. $O_{1,i}$ represents the output of the i^{th} node, and μ_{Ai} is the MF of Ai. Based on the literature the ANFIS developed here utilizes Gaussian MF due to its superior performance (Abd El Aziz et al., 2017). For the generalized Gaussian MF, ρ and σ are the mean and SD of the dataset, respectively. The ANFIS used seven MFs and 500 training epochs to give the optimal results, which is implemented in MATLAB R2021b.

2.3. Multi-layer Perceptron Neural Network

MLPs are fully connected feedforward NN used in supervised regression problems and utilize Levenberg-Marquardt back-propagation algorithm for training. The input data was applied to the MLP-NN in this training process, and the output of network output was calculated employing randomly chosen initial weights. The output was then compared with the target and a weight correction process was performed in the opposite direction of the mean square error gradient. As a result, the difference between the network output and the desired output could be reduced.

MLP-NN generally has three layers: input layer, hidden layer, and output layer. Typically, it consists of small processing units called neurons. Each neuron composes of five components: input layer, basis and bias, summation function, activation function, and target. In this work the MLP-NN architecture has an input, an output layer, and the number of hidden layers varies between one, two, and three (Ascher et al., 2022). Previously, researchers have shown that this architecture can capture complex non-linear characteristic (Asgari et al., 2021b; Asgari et al., 2021c). A representative schematic of the MLP-NN is shown in Fig. 2.

The activation function, also known as the transfer function, is either a linear or non-linear function that converts the weighted sum of the inputs (internally generated sum) into an output value. Here, three different activation functions such as (a) rectified linear unit (ReLU), (b) hyperbolic tangent (Tanh), and (c) Sigmoid (see Eqs. ((4)–(6))) were considered. Different hyperparameters for the MLP-NN were also explored where the number of neurons varied from 8 to 16 and number of hidden layers were varied from 1 to 3. Three different combinations of data splitting for model training and testing were also evaluated i.e., 80%/20%, 70%/30%, and 60%/40%. Other fixed parameters for the MLP-NN are maximum epoch = 300 and learning rate = 0.01. The entire model was implemented in Anaconda (a Python environment) with aid of the Sklearn library.

$$\mathrm{ReLU} = \max(0, x) \tag{4}$$

$$\mathrm{Tanh} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{5}$$

$$\mathrm{Sigmoid} = \frac{1}{1 + e^{-x}} \tag{6}$$

2.4. Model accuracy evaluation metrics

The performance of two data-driven models is examined and evaluated using different training–testing split on the dataset containing 226 datapoints. Two different metrics are considered, which are popular for ML regression problems such as *RMSE* and R^2, calculated as,

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

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

Here Y_i^{exp} and Y_i^{pred} are the experimental and model predicted biochar yields, Y_{ave}^{exp} is the average of all the experimental biochar yields, and N is the total number of datapoints which is 226 in this work.

3. Results and discussion

3.1. Exploration of dataset

The statistical characteristics for input (FC_{FS}, VM_{FS}, ash_{FS}, C_{FS}, H_{FS}, O_{FS}, N_{FS}, PT, HR, and RT) and output ($Yield_{BC}$, FC_{BC}, FC_{BC}, ash_{BC}, C_{BC}, H_{BC}, O_{BC}, and N_{BC}) variables were quantified via the means and SDs of the input variables as shown in Table 1. Furthermore, the linear correlation among any two variables was assessed by *PCC* (see Eq.3) and presented in form of a heatmap (Fig. 3a). Here, $PCC \approx 0$ signifies that the variables are weakly correlated, whereas $PCC \approx \pm 1$ suggests the highest correlation strength. Among the input parameters, strong correlations (with $PCC \geq 0.4$) were observed between several components of proximate and ultimate composition of biomass feedstock ($PCC = 0.71$ FC_{FS} vs. VM_{FS}, $PCC = -0.54$ for C_{FS} vs. ash_{FS}, $PCC = 0.48$ for FC_{FS} vs. ash_{FS}, $PCC = 0.41$ for H_{FS} vs. O_{FS}). Similar works in the literature reported the existence of a correlation between the proximate and ultimate compositions of biomass feedstock (Khan et al., 2022; Zhu et al., 2019).

Subsequently, the $|PCC|$ values among inputs and outputs were visualized in Fig. 3b to 3d, which signify the relative importance of an input feature in predicting outputs. Fig. 3b revealed that the $Yield_{BC}$ is

Table 2
Simplified formulation of the ANFIS framework implemented in MATLAB.

Attribute	Mathematical structure
General rule	Rule 1: IF X is $A1$ AND Y is $B1$, THEN $f_1 = p_1X + q_1Y + r_1$ Rule 2: IF X is $A2$ AND Y is $B2$, THEN $f_2 = p_2X + q_2Y + r_2$
First layer	$O_{1,i} = \mu_{Ai}(X), i = 1,2$ $\mu_{Ai}(X) = e^{\frac{-(X-\rho)^2}{2\sigma^2}}, i = 1,2$
Second layer	$O_{2,i} = W_i = \mu_{Ai}(X) \times \mu_{Bi}(Y), i = 1,2$
Third layer	$O_{3,i} = \overline{W_i} = \frac{W_i}{W_1 + W_2}, i = 1,2$
Fourth layer	$O_{4,i} = \overline{W_i} f_i = \overline{W_i}(p_iX + q_iY + r_i), i = 1,2$
Fifth layer	$O_5 = \sum_i \overline{W_i} f_i = \text{final output}$

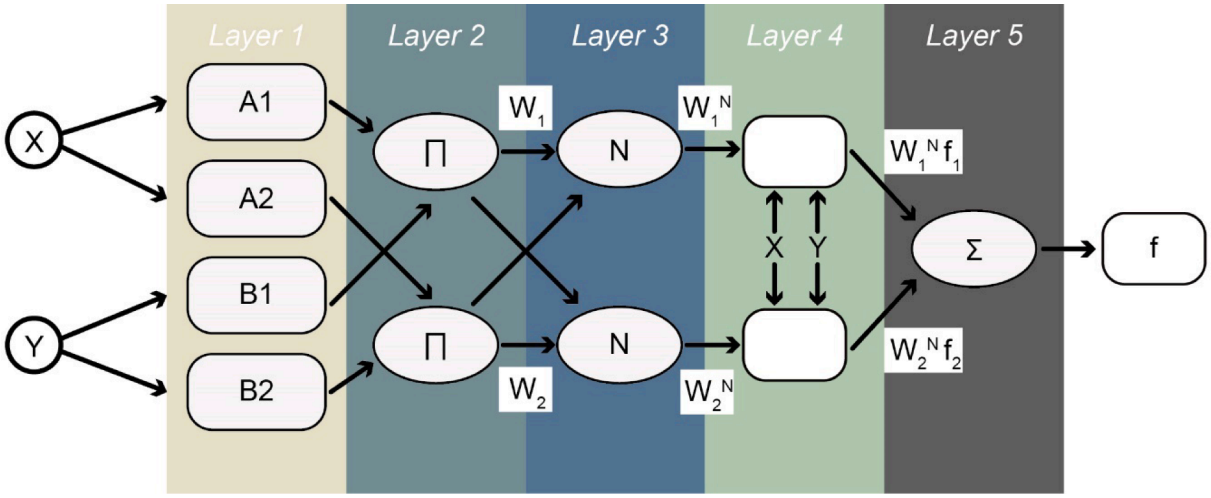


Fig. 1. Schematic representation of the ANFIS model.

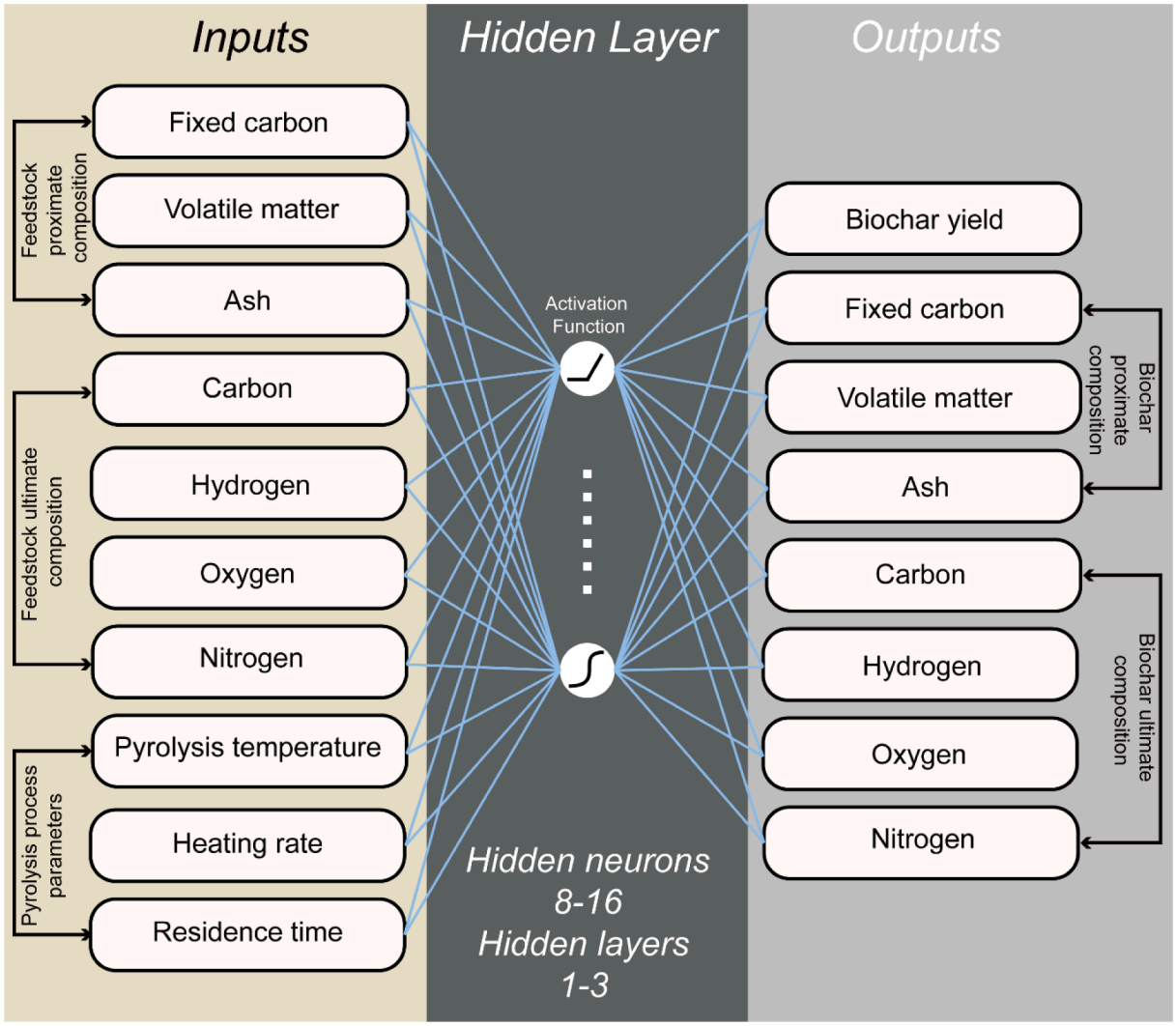


Fig. 2. Schematic representation of the MLP-NN architecture showing input variables, hidden layers, and output variables.

strongly affected by the input variable PT ($|PCC| = 0.76$). The negative symbol of *PCC* for PT in Fig. 3a signified that increasing PT will reduce biochar yield. However, increasing PT could improve the carbon content in the biochar by reducing volatile components such as H, O, and N in biochar, which indicated the existence of a tradeoff (Marx et al., 2014). Fig. 3c showed the dependence of biochar proximate compositions (FC_{BC}, FC_{BC}, and ash_{BC}) on input variables. Strong correlations were observed between the following output vs. input pairs (in decreasing

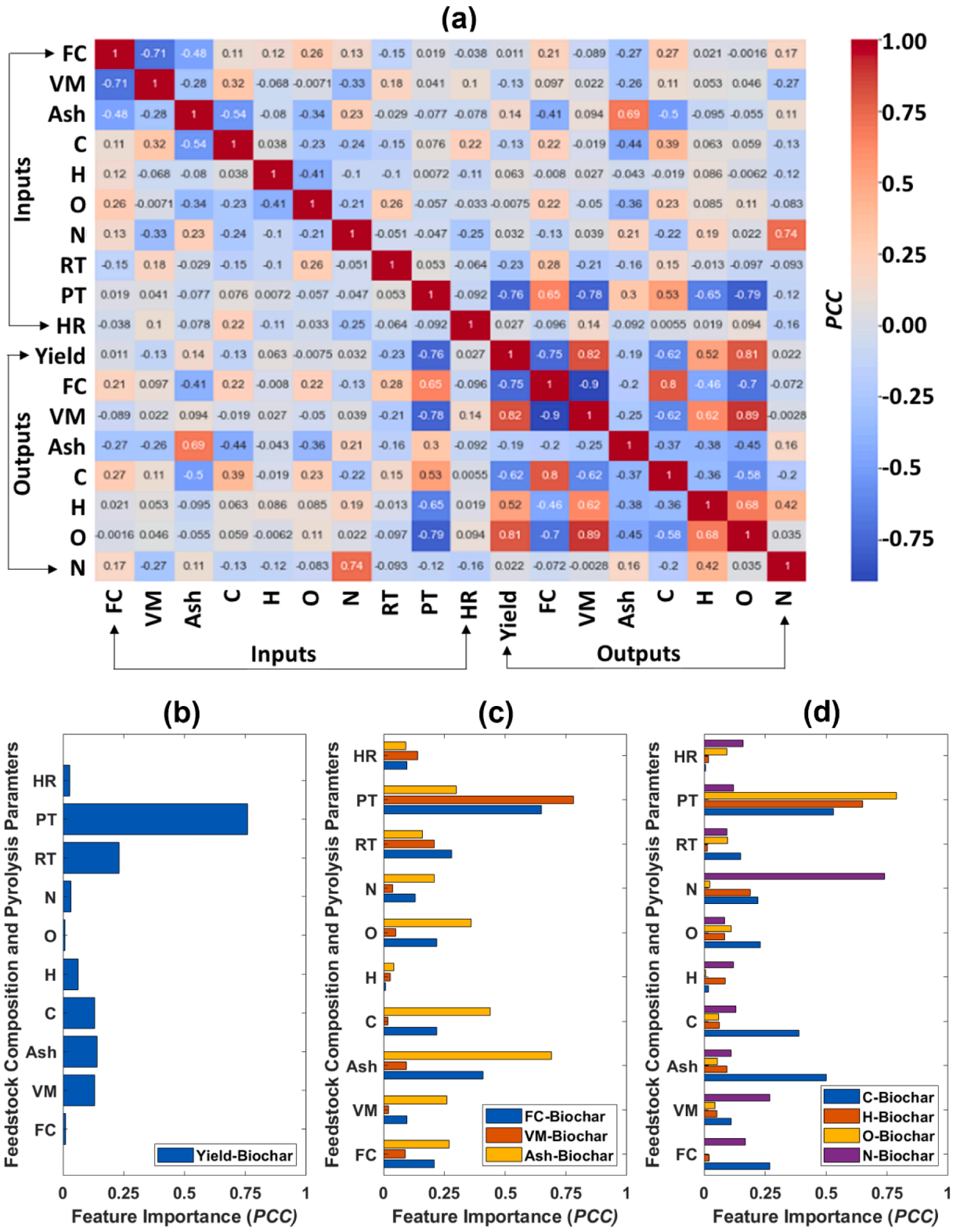


Fig. 3. (a) *PCC* between any two variables of interest. Relative importance of input features for predicting outputs: (b) biochar yield, (c) proximate composition of biochar (FC-VM-ash) and (d) ultimate composition of biochar (C-H-O-N).

order): $|PCC| = 0.78$ for FC_{BC} vs. PT, $|PCC| = 0.65$ for ash_{BC} vs. ash_{FS}, $|PCC| = 0.65$ for FC_{BC} vs. PT, and $|PCC| = 0.41$ for FC_{BC} vs. ash_{FS}. Similarly, Fig. 3d revealed the significant dependence of biochar ultimate composition ($C_{BC}, H_{BC}, O_{BC}, N_{BC}$) on input variables as follows (in decreasing order): $|PCC| = 0.79$ for O_{BC} vs. PT, $|PCC| = 0.74$ for N_{BC} vs. N_{FS}, $|PCC| = 0.65$ for H_{BC} vs. PT, $|PCC| = 0.53$ for C_{BC} vs. PT, and $|PCC| = 0.5$ for C_{BC} vs. ash_{FS}. Exploration of the feature importance demonstrated the strong influence of PT, affecting 6 out of 8 output variables including $Yield_{BC}$. The remaining two variables ash_{BC} and N_{BC} were affected by ash_{FS} and N_{FS}, respectively. Therefore, any uncertainties associated with these variables would strongly affect the predictive performance of the data-driven models.

3.2. Predictive performance of Multi-layer Perceptron Neural network

The predictive accuracy of an MLP-NN model depends on the hyperparameters selection, which includes choices for (1) the number of neurons within a hidden layer, (2) the number of hidden layers, and (3) the type of activation function. Table 3 lists various scenarios selected for determining the optimal hyperparameters. Cases 1A to 1E shows the influence of varying the number of neurons from 8 to 16 in an interval of 2. The fixed parameters for Cases 1A to 1E are 1 hidden layer, ReLU as the activation function, 80% training data, and 20% testing data. The choice of training–testing split is consistent with a prior work related to data-driven biochar yield prediction (Zhu et al., 2019). It is observed that as the number of neurons increases, the model accuracy improves at first. Beyond a certain value of neurons, i.e., 14 in this case, the model accuracy degrades. When the MLP-NN consists of a lower number of neurons under a critical value (14 neurons), the model underfits the data and therefore results in higher *RMSE* and poor R^2. However, when the number of neurons is increased beyond a critical value, the MLP-NN model overfits and fails to achieve a higher accuracy in predicting the testing dataset (Asgari et al., 2021a; Asgari et al., 2021c). Overall, the optimal outcome of Cases 1A to 1E has the following accuracies for predicting biochar yield ($R^2 = 0.964$, $RMSE = 3.4$), proximate composition (average $R^2 = 0.914$, average $RMSE = 4.4$), and ultimate composition (average $R^2 = 0.887$, average $RMSE = 1.8$).

Once the number of neurons is fixed to 14 for the MLP-NN model, it is essential to further explore if increasing the number of hidden layers can improve the model predictions. Cases 2A to 2C in Table 3 correspond to increasing the hidden layer from 1 to 3. The fixed parameters for these cases are 14 neurons within a layer, ReLU as the activation function, 80% training data, and 20% testing data. Comparing the cases with the different number of hidden layers reveals that increasing the number of hidden layers to 2 and 3 diminishes the model prediction accuracy by up to 2.2% and 4.9% (with respect to Case 2A), respectively. Hence, using a single hidden layer in this case will not cause model overfitting and lead to better accuracy than a denser MLP-NN (Asgari et al., 2021c).

Subsequently, different activation functions such as ReLU (Case 3A), Tanh (Case 3B), and Sigmoid (Case 3C) were tested by fixing the number of neurons to 14 and number of hidden layers to 1. Utilizing the Tanh and Sigmoid activation functions degrade the model performance by up to 4.8% and 9.6%, respectively, when compared to the ReLU activation function. The nature of the ReLU activation function (see Eq. (4)) avoids the vanishing gradient while training the MLP-NN model, since the function does not have any hard-bound upper limit. The Tanh activation function has a range of −1 to 1 (Eq. (5)), while the range for the Sigmoid activation function is from 0 to 1 (Eq. (6)). Therefore, the Sigmoid activation function is the most vulnerable to the vanishing gradient problem due to its shortest range, decreasing the accuracy of the model. It is essential to note that the MLP-NN model with both ReLU and Tanh activation functions significantly outperforms the existing works on biochar yield prediction which reports R^2 values 0.84 (Pathy et al., 2020) and 0.85 (Zhu et al., 2019) using the XGB and RF models, respectively. Moreover, the model developed had 3.7% better predictive accuracy, when compared to the best model reported ($R^2 = 0.93$) in the biochar yield prediction literature (Khan et al., 2022). The parity plots shown in Fig. 4 compared the predicted and actual values for 8 different outputs for the optimal scenario of MLP-NN.

3.3. Predictive performance of artificial Neuro-Fuzzy inference system

The predictive accuracy of ANFIS developed in this work was shown via the parity plots in Fig. 5. Overall, the ANFIS model had the following prediction accuracies for biochar yield ($R^2 = 0.877$, $RMSE = 4.9$), proximate composition (average $R^2 = 0.838$, average $RMSE = 5.9$), and ultimate composition (average $R^2 = 0.855$, average $RMSE = 2.2$). The ANFIS model with Gaussian MF, 7 MFs per input, 80% training data, and 20% testing data provided the optimal results. It was also found that the predictive accuracy of ANFIS for biochar yield, proximate composition, and ultimate composition were 9%, 8.3%, and 3.6% lower than those for MLP-NN. Nevertheless, the model provided a competitive R^2 value for predicting biochar yield compared to those reported in the literature (Pathy et al., 2020; Zhu et al., 2019).

3.4. Comparison of MLP-NN, ANFIS, and existing works

The generalizability of the MLP-NN and ANFIS model for predicting biochar yield, proximate composition, and ultimate composition was further investigated by varying the percentage of training testing dataset split. Three different training/testing percentage splits were considered for the overall database (226 datasets): 80%/20%, 70%/30%, and 60%/40%. Table 4 compares (1) MLP-NN with ReLu (Cases 1–3), (2) MLP-NN

Table 3
Predictive performance of MLP-NN for variations in number of neurons, number hidden layers, and types of activation function for 80% training data and 20% testing data. R^2 and *RMSE* values for biochar yield, proximate composition (FC-VM-ash), and ultimate composition (C-H-O-N) are shown. Descriptions of various cases are described the text.

Case	Attribute	$Yield_{BC}$	FC_{BC}	VM_{BC}	ash_{BC}	C_{BC}	H_{BC}	O_{BC}	N_{BC}
1A	R^2	0.861	0.859	0.835	0.853	0.885	0.795	0.813	0.858
	RMSE	6.5	6.7	7.7	3.4	3.7	0.9	4.1	0.5
1B	R^2	0.892	0.880	0.861	0.874	0.893	0.823	0.846	0.876
	RMSE	5.1	6.2	6.9	2.9	3.6	0.7	3.8	0.4
1C	R^2	0.927	0.891	0.881	0.934	0.909	0.849	0.880	0.872
	RMSE	4.0	5.8	6.4	2.1	3.3	0.5	3.4	0.4
1D	R^2	0.964	0.903	0.899	0.940	0.918	0.855	0.887	0.888
	RMSE	3.4	5.1	6.1	2.0	3.1	0.5	3.3	0.4
1E	R^2	0.940	0.883	0.879	0.921	0.911	0.839	0.884	0.865
	RMSE	3.8	6.1	6.5	2.4	3.2	0.6	3.3	0.5
2A	R^2	0.964	0.903	0.899	0.940	0.918	0.855	0.887	0.888
	RMSE	3.4	5.1	6.1	2.0	3.1	0.5	3.3	0.4
2B	R^2	0.943	0.896	0.882	0.927	0.910	0.848	0.872	0.874
	RMSE	3.7	5.3	6.4	2.4	3.3	0.5	3.5	0.4
2C	R^2	0.917	0.879	0.865	0.918	0.901	0.839	0.853	0.871
	RMSE	4.2	6.2	6.8	2.5	3.5	0.6	3.7	0.4
3A	R^2	0.964	0.903	0.899	0.940	0.918	0.855	0.887	0.888
	RMSE	3.4	5.1	6.1	2.0	3.1	0.5	3.3	0.4
3B	R^2	0.918	0.888	0.891	0.929	0.901	0.846	0.869	0.876
	RMSE	4.2	5.8	6.1	2.3	3.3	0.5	3.6	0.4
3C	R^2	0.871	0.846	0.862	0.889	0.874	0.821	0.836	0.851
	RMSE	5.9	6.9	6.9	2.6	3.9	0.7	3.9	0.6

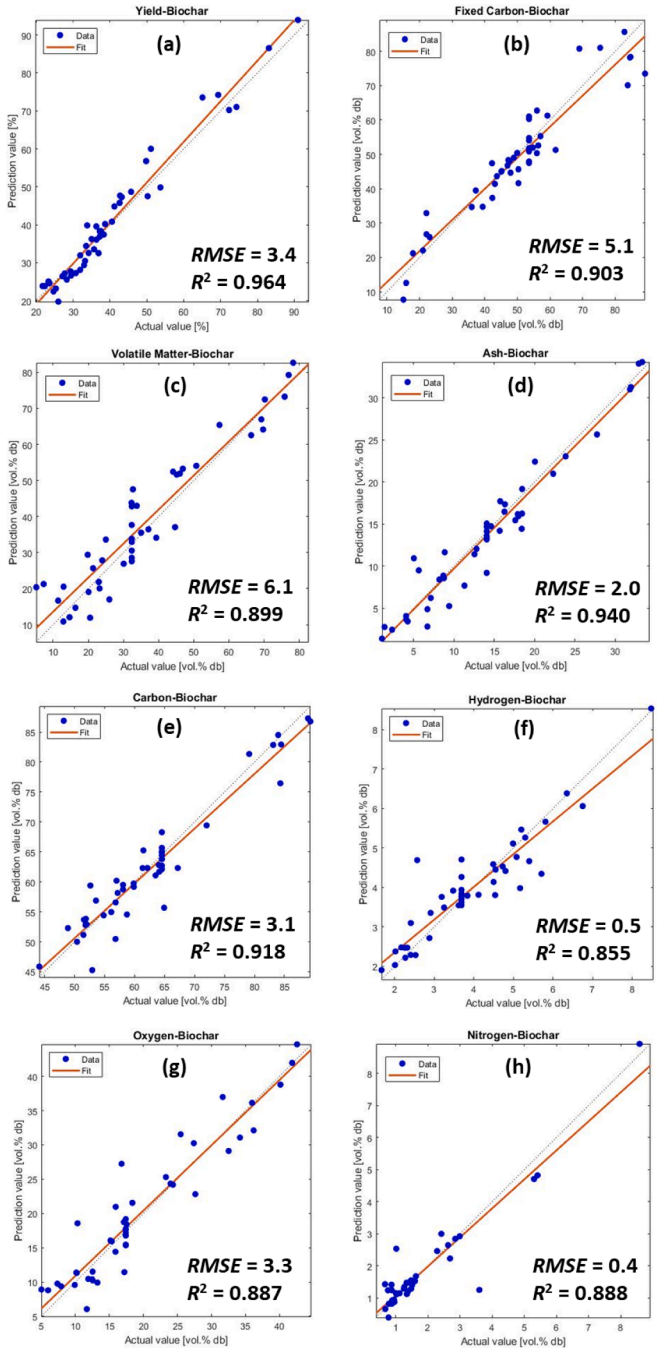


Fig. 4. Parity plots for MLP-NN comparing the actual and predicted values of biochar (a) yield, (b-d) proximate composition (FC-VM-ash) and (e-h) ultimate composition (C-H-O-N). The parity plots correspond to the optimal model training scenario with 14 neurons, 1 hidden layer, ReLU activation function, 80%/20% training–testing data split.

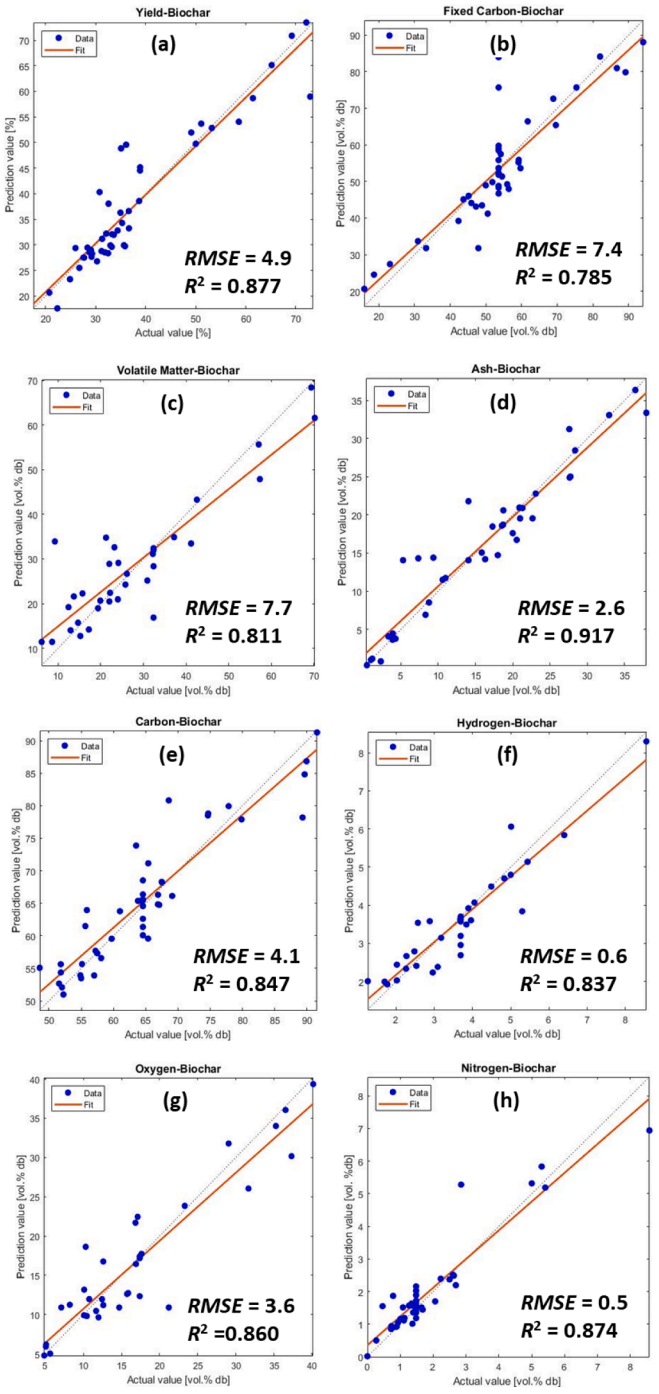


Fig. 5. Parity plots for ANFIS comparing the actual and predicted values of biochar (a) yield, (b-d) proximate composition (FC-VM-ash) and (e-h) ultimate composition (C–H-O-N). The parity plots correspond to the Gaussian membership function, 7 membership functions per unit, and 80%/20% training–testing data split.

with Tanh (Cases 4–6), and (3) ANFIS (cases 7–9) for different training/testing dataset split percentages. These results were further compared to relevant works from the data-driven biochar yield prediction literature (Cao et al., 2016; Khan et al., 2022; Pathy et al., 2020; Zhu et al., 2019).

Examining cases 1–3 revealed that changing the training–testing split from 80%/20% to 70%/30% decreases the predictive accuracy for biochar yield by 3.5%, proximate composition by 2.5%, and ultimate composition 2.2%. This suggested that with a 10% smaller training dataset, the performance of MLP-NN with ReLU was not significantly affected. Similar assessment was done for MLP-NN with Tanh activation function by comparing Cases 4–6. In this scenario, changing the training–testing split proportions from 80%/20% to 70%/30% degraded the predictive accuracy for biochar yield by 3.8%, proximate composition by 2.7%, and ultimate composition by 2.3%. For ANFIS (Cases 7–9), similar analysis reveals a performance degradation of 0.7% for biochar yield, 1.1% for proximate composition, and 2.1% for ultimate composition. Overall, the predictive accuracy of the MLP-NN with ReLU is up to 9.9% time better than that for ANFIS. An additional incentive offered by the MLP-NN model is that its computation time is 8 min for the present work, which is approximately 3 times faster than the ANFIS

Table 4
R^2 and *RMSE* values for biochar yield, proximate composition (FC-VM-ash), and ultimate composition (C-H-O-N) prediction using different proportions of training and testing data. The results are also compared to those reported in the biochar yield prediction literature. Descriptions of cases A to I are described the text.

Case	Attribute	$Yield_{BC}$	FC_{BC}	VM_{BC}	ash_{BC}	C_{BC}	H_{BC}	O_{BC}	N_{BC}
A	R^2	0.964	0.903	0.899	0.940	0.918	0.855	0.887	0.888
	RMSE	3.4	5.1	6.1	2.0	3.1	0.5	3.3	0.4
B	R^2	0.930	0.876	0.871	0.926	0.896	0.841	0.863	0.871
	RMSE	3.9	6.1	6.6	2.4	3.6	0.6	3.6	0.4
C	R^2	0.891	0.827	0.820	0.886	0.851	0.826	0.829	0.856
	RMSE	5.1	7.5	8.1	2.9	4.4	0.7	5.4	0.5
D	R^2	0.918	0.888	0.891	0.929	0.901	0.846	0.869	0.876
	RMSE	4.2	5.8	6.1	2.3	3.3	0.5	3.6	0.4
E	R^2	0.883	0.861	0.862	0.911	0.878	0.832	0.841	0.860
	RMSE	5.2	6.6	6.8	2.5	3.9	0.6	4.2	0.5
F	R^2	0.851	0.821	0.811	0.879	0.829	0.815	0.819	0.848
	RMSE	6.6	7.6	8.3	3.1	5.0	0.8	5.9	0.6
G	R^2	0.877	0.785	0.811	0.917	0.847	0.837	0.860	0.874
	RMSE	4.9	7.4	7.7	2.6	4.1	0.6	3.6	0.5
H	R^2	0.871	0.777	0.798	0.909	0.821	0.826	0.839	0.859
	RMSE	4.9	7.5	7.9	2.7	4.8	0.6	4.3	0.6
I	R^2	0.862	0.756	0.773	0.887	0.802	0.809	0.796	0.839
	RMSE	5.6	7.7	8.1	3.1	5.2	0.8	6.1	0.7
RF model (Zhu et al., 2019)	R^2	0.855	—	—	—	0.848	—	—	—
	RMSE	3.4	—	—	—	5.8	—	—	—
SVM model (Cao et al., 2016)	R^2	0.804	—	—	—	—	—	—	—
	RMSE	6.4	—	—	—	—	—	—	—
XGB model (Pathy et al., 2020)	R^2	0.844	—	—	—	—	—	—	—
	RMSE	—	—	—	—	—	—	—	—
ANN model (Khan et al., 2022)	R^2	0.93	—	—	—	—	—	—	—
	RMSE	1.74	—	—	—	—	—	—	—

model. All the model trainings were run on a desktop with 11th Gen Intel Core i5-11400F @ 2.50 GHz processor with 8 GB RAM.

Further, the MLP-NN with ReLU activation function was compared to relevant works in the biochar prediction literature. It is important to note that relevant work in the biochar prediction literature is mostly limited to biochar yield prediction (Cao et al., 2016; Khan et al., 2022; Pathy et al., 2020), except one where the carbon content of biochar was predicted in parallel to yield (Zhu et al., 2019). The MLP-NN with ReLU model has 4.8% better prediction accuracy than the best possible (R^2 = 0.92) biochar yield prediction model (Khan et al., 2022). Simultaneously, the MLP-NN model has 8.2% better predictive accuracy for predicting carbon contents in biochar when compared to the RF model (Zhu et al., 2019). In addition, the present work predicts other elemental compositions (H-O-N) and proximate compositions (FC-VM-ash) of biochar produced via pyrolysis. These additional advantages proved the superiority of the present work, where a comprehensive model was developed for the biochar production process. Future research efforts are recommended to study the comparison of various data-driven models such as RF, XGB, Support Vector Machine (SVM), and MLP-NN subjected to the same input dataset for simultaneous prediction of biochar yield and compositions.

4. Conclusions

Comprehensive data-driven models were developed to predict biochar yield and compositions based on pyrolysis conditions and biomass feedstock compositions. Feature importance analysis revealed high dependence of biochar yield and composition on pyrolysis temperature, ash content, and amount of nitrogen. The MLP-NN had predictive accuracies for biochar yield (R^2 = 0.964, *RMSE* = 3.4), proximate composition (R^2 = 0.914, *RMSE* = 4.4), and ultimate composition (R^2 = 0.887, *RMSE* = 1.8). The MLP-NN showed 9.9% and 12.7% performance improvement than ANFIS and existing works in the literature, respectively. This opens avenues for future research in data-driven biochar process modelling and optimization.

CRediT authorship contribution statement

Yize Li: Conceptualization, Data curation, Investigation, Methodology, Software, Visualization, Writing – original draft. **Rohit Gupta:** Conceptualization, Investigation, Methodology, Supervision, Visualization, Writing – original draft. **Siming You:** Conceptualization, Funding acquisition, Methodology, Project administration, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All the associated data have been provided in the manuscript and electronic supplementary material.

Acknowledgments

The authors would like to acknowledge the financial support from the UK Engineering and Physical Sciences Research Council (EPSRC) Programme Grant (EP/V030515/1). Siming You would like to thank the financial support from the Royal Society Research Grant (RGS\R1 \211358) and International Exchange Scheme (EC\NSFC\211175). The authors would also like to acknowledge fruitful discussions with Mr. Simon Ascher from the University of Glasgow towards model improvement. All data supporting this study are provided in full in the 'Methodology' and 'Results and Discussion' sections, and supplementary information of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biortech.2022.127511.

References

Abd El Aziz, M., Hemdan, A.M., Ewees, A.A., Elhoseny, M., Shehab, A., Hassanien, A.E., Xiong, S. 2017. Prediction of biochar yield using adaptive neuro-fuzzy inference system with particle swarm optimization. *2017 IEEE PES PowerAfrica*. IEEE. pp. 115-120.

Ascher, S., Sloan, W., Watson, I., You, S., 2022. A comprehensive artificial neural network model for gasification process prediction. Appl. Energy 320, 119289.

Asgari, S., Gupta, R., Puri, I.K., Zheng, R., 2021a. A data-driven approach to simultaneous fault detection and diagnosis in data centers. Appl. Soft Comput. 110, 107638.

Asgari, S., MirhoseiniNejad, S., Moazamigoodarzi, H., Gupta, R., Zheng, R., Puri, I.K., 2021b. A gray-box model for real-time transient temperature predictions in data centers. Appl. Therm. Eng. 185, 116319.

Asgari, S., Moazamigoodarzi, H., Tsai, P.J., Pal, S., Zheng, R., Badawy, G., Puri, I.K., 2021c. Hybrid surrogate model for online temperature and pressure predictions in data centers. Fut. Generat. Comput. Syst. 114, 531–547.

Bakyani, A.E., Sahebi, H., Ghiasi, M.M., Mirjordavi, N., Esmaeilzadeh, F., Lee, M., Bahadori, A., 2016. Prediction of CO_2–oil molecular diffusion using adaptive neuro-fuzzy inference system and particle swarm optimization technique. Fuel 181, 178–187.

Bhattacharjee, N., Biswas, A.B., 2019. Pyrolysis of orange bagasse: comparative study and parametric influence on the product yield and their characterization. J. Environ. Chem. Eng. 7 (1), 102903.

Biswas, B., Pandey, N., Bisht, Y., Singh, R., Kumar, J., Bhaskar, T., 2017. Pyrolysis of agricultural biomass residues: Comparative study of corn cob, wheat straw, rice straw and rice husk. Bioresour. Technol. 237, 57–63.

Bolan, N., Hoang, S.A., Beiyuan, J., Gupta, S., Hou, D., Karakoti, A., Joseph, S., Jung, S., Kim, K.-H., Kirkham, M., 2021. Multifunctional applications of biochar beyond carbon storage. Int. Mater. Rev. 1–51.

Brassard, P., Godbout, S., Pelletier, F., Raghavan, V., Palacios, J.H., 2018. Pyrolysis of switchgrass in an auger reactor for biochar production: a greenhouse gas and energy impacts assessment. Biomass Bioenergy 116, 99–105.

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

Cha, J.S., Park, S.H., Jung, S.-C., Ryu, C., Jeon, J.-K., Shin, M.-C., Park, Y.-K., 2016. Production and utilization of biochar: a review. J. Ind. Eng. Chem. 40, 1–15.

Chen, D., Liu, D., Zhang, H., Chen, Y., Li, Q., 2015. Bamboo pyrolysis using TG–FTIR and a lab-scale reactor: analysis of pyrolysis behavior, product properties, and carbon and energy yields. Fuel 148, 79–86.

Crombie, K., Mašek, O., 2015. Pyrolysis biochar systems, balance between bioenergy and carbon sequestration. GCB Bioenergy 7 (2), 349–361.

Crombie, K., Mašek, O., Sohi, S.P., Brownsort, P., Cross, A., 2013. The effect of pyrolysis conditions on biochar stability as determined by three methods. GCB Bioenergy 5 (2), 122–131.

Duan, M., Wu, F., Jia, Z., Wang, S., Cai, Y., Chang, S.X., 2020. Wheat straw and its biochar differently affect soil properties and field-based greenhouse gas emission in a Chernozemic soil. Biol. Fertil. Soils 56 (7), 1023–1036.

Gwenzi, W., Chaukura, N., Wenga, T., Mtisi, M., 2021. Biochars as media for air pollution control systems: contaminant removal, applications and future research directions. Sci. Total Environ. 753, 142249.

He, X., Liu, Z., Niu, W., Yang, L., Zhou, T., Qin, D., Niu, Z., Yuan, Q., 2018. Effects of pyrolysis temperature on the physicochemical properties of gas and biochar obtained from pyrolysis of crop residues. Energy 143, 746–756.

Hong, Z., Zhong, F., Niu, W., Zhang, K., Su, J., Liu, J., Li, L., Wu, F., 2020. Effects of temperature and particle size on the compositions, energy conversions and structural characteristics of pyrolysis products from different crop residues. Energy 190, 116413.

Irfan, M., Chen, Q., Yue, Y., Pang, R., Lin, Q., Zhao, X., Chen, H., 2016. Co-production of biochar, bio-oil and syngas from halophyte grass (Achnatherum splendens L.) under three different pyrolysis temperatures. Bioresour. Technol. 211, 457–463.

Kaczor, Z., Buliński, Z., Werle, S., 2020. Modelling approaches to waste biomass pyrolysis: a review. Renewable Energy 159, 427–443.

Khan, M., Ullah, Z., Mašek, O., Naqvi, S.R., Khan, M.N.A., 2022. Artificial neural networks for the prediction of biochar yield: a comparative study of metaheuristic algorithms. Bioresour. Technol. 355, 127215.

Lee, J., Kim, K.-H., Kwon, E.E., 2017. Biochar as a catalyst. Renew. Sustain. Energy Rev. 77, 70–79.

Lee, Y., Park, J., Ryu, C., Gang, K.S., Yang, W., Park, Y.-K., Jung, J., Hyun, S., 2013. Comparison of biochar properties from biomass residues produced by slow pyrolysis at 500 C. Bioresour. Technol. 148, 196–201.

Li, Y., Ahmed, A., Watson, I., You, S. 2020. Waste-to-biofuel and carbon footprints. in: *Waste Biorefinery*, Elsevier, pp. 579-597.

Liu, X., Zhang, Y., Li, Z., Feng, R., Zhang, Y., 2014. Characterization of corncob-derived biochar and pyrolysis kinetics in comparison with corn stalk and sawdust. Bioresour. Technol. 170, 76–82.

Liu, Z., Han, G., 2015. Production of solid fuel biochar from waste biomass by low temperature pyrolysis. Fuel 158, 159–165.

Liu, Z., Niu, W., Chu, H., Niu, Z., 2018a. Process optimization for straws pyrolysis and analysis of biochar physiochemical properties. Trans. Chin. Soc. Agric. Eng. 34 (5), 196–203.

Liu, Z., Niu, W., Chu, H., Zhou, T., Niu, Z., 2018b. Effect of the carbonization temperature on the properties of biochar produced from the pyrolysis of crop residues. BioResources 13 (2), 3429–3446.

Liu, Z., Zhang, F., Liu, H., Ba, F., Yan, S., Hu, J., 2018c. Pyrolysis/gasification of pine sawdust biomass briquettes under carbon dioxide atmosphere: study on carbon dioxide reduction (utilization) and biochar briquettes physicochemical properties. Bioresour. Technol. 249, 983–991.

Marx, S., Chiyanzu, I., Piyo, N., 2014. Influence of reaction atmosphere and solvent on biochar yield and characteristics. Bioresour. Technol. 164, 177–183.

Pathy, A., Meher, S., Balasubramanian, P., 2020. Predicting algal biochar yield using eXtreme Gradient Boosting (XGB) algorithm of machine learning methods. Algal Res. 50, 102006.

Patra, B.R., Nanda, S., Dalai, A.K., Meda, V., 2021. Slow pyrolysis of agro-food wastes and physicochemical characterization of biofuel products. Chemosphere 285, 131431.

Qiu, B., Tao, X., Wang, H., Li, W., Ding, X., Chu, H., 2021. Biochar as a low-cost adsorbent for aqueous heavy metal removal: a review. J. Anal. Appl. Pyrol. 155, 105081.

Rout, T., Pradhan, D., Singh, R., Kumari, N., 2016. Exhaustive study of products obtained from coconut shell pyrolysis. J. Environ. Chem. Eng. 4 (3), 3696–3705.

Sajjadi, B., Chen, W.-Y., Egiebor, N.O., 2019. A comprehensive review on physical activation of biochar for energy and environmental applications. Rev. Chem. Eng. 35 (6), 735–776.

Shariff, A., Noor, N.M., Lau, A., Ali, M.A.M., 2016. A comparative study on biochar from slow pyrolysis of corn cob and cassava wastes. Int. J. Biotechnol. Bioeng. 10 (12), 767–771.

Tag, A.T., Duman, G., Ucar, S., Yanik, J., 2016. Effects of feedstock type and pyrolysis temperature on potential applications of biochar. J. Anal. Appl. Pyrol. 120, 200–206.

Tripathi, N., Hills, C.D., Singh, R.S., Atkinson, C.J., 2019. Biomass waste utilisation in low-carbon products: harnessing a major potential resource. NPJ Clim. Atmos. Sci. 2 (1), 1–10.

Ucar, S., Ozkan, A.R., 2008. Characterization of products from the pyrolysis of rapeseed oil cake. Bioresour. Technol. 99 (18), 8771–8776.

Wang, Z., Peng, X., Xia, A., Shah, A.A., Huang, Y., Zhu, X., Zhu, X., Liao, Q., 2022. The role of machine learning to boost the bioenergy and biofuels conversion. Bioresour. Technol. 343, 126099.

Zhang, Y., Ma, Z., Zhang, Q., Wang, J., Ma, Q., Yang, Y., Luo, X., Zhang, W., 2017. Comparison of the physicochemical characteristics of bio-char pyrolyzed from moso bamboo and rice husk with different pyrolysis temperatures. BioResources 12 (3), 4652–4669.

Zhu, X., Li, Y., Wang, X., 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. 288, 121527.