

# Artificial Neural Network Modeling to Predict the Efficiency of Phosphoric Acid-Hydrogen Peroxide Pretreatment of Wheat Straw

Qing Wang,<sup>a,b,c</sup> Jinxiang Hua,<sup>b</sup> Jinguang Hu,<sup>c</sup> Li Zhao,<sup>b</sup> Mei Huang,<sup>b</sup> Dong Tian,<sup>b</sup> Yongmei Zeng,<sup>b</sup> Shihuai Deng,<sup>b</sup> Fei Shen,<sup>b,\*</sup> and Xinquan Zhang<sup>a,\*</sup>

Phosphoric acid-hydrogen peroxide (PHP) pretreatment is an effective method to obtain a cellulose-enriched fraction from biomass. In this study, artificial neural network (ANN) was used to predict PHP pretreatment efficiency of cellulose content (C-C), cellulose recovery (C-R<sub>y</sub>), hemicellulose removal (H-R<sub>i</sub>), and lignin removal (L-R<sub>i</sub>) under various conditions of pretreatment time (*t*), temperature (*T*), H<sub>3</sub>PO<sub>4</sub> concentration (C<sub>p</sub>), and H<sub>2</sub>O<sub>2</sub> concentration (C<sub>h</sub>). The final optimized topology structure of the ANN models had 1 hidden layers with 9 neurons for C-C and 10 neurons for C-R<sub>y</sub>, 10 neurons for H-R<sub>i</sub>, and 12 neurons for L-R<sub>i</sub>. The actual testing data fit the predicted data with R<sup>2</sup> values ranging from 0.8070 to 0.9989. The relative importance (RI) revealed that C<sub>p</sub> and C<sub>h</sub> were significant factors influencing the efficiency of PHP pretreatment with total RI values ranging from 12% to 62.6%. However, their weights for the three components of biomass were different. The value of *T* dominated hemicellulose removal effectiveness with an RI value of 78.6%, while *t* did not seem to be a main factor dominating PHP pretreatment efficiency. The results of this study provide insights into the convenient development and optimization of biomass pretreatment from ANN modeling perspectives.

DOI: 10.15376/biores.19.1.288-305

Keywords: Lignocellulosic biomass; ANN model; Pretreatment efficiency; Prediction; Relative importance

Contact information: a: College of Grassland Science and Technology, Sichuan Agricultural University, Chengdu, Sichuan 611130, P. R. China; b: Institute of Ecological and Environmental Sciences, Sichuan Agricultural University, Chengdu, Sichuan 611130, P. R. China; c: Department of Chemical and Petroleum Engineering, Schulich School of Engineering, the University of Calgary, Calgary T2N 4H9, Canada; \*Corresponding authors: fshen@sicau.edu.cn; zhangxq@sicau.edu.cn

## INTRODUCTION

The rapid expansion of human society relies on the consumption of fossil resources, such as petroleum, coal, and natural gas, which cause severe problems, including environmental contamination and climate change (Rashid *et al.* 2021). This situation has led to an increased interest in alternative and renewable energy sources (Hosseini *et al.* 2019). Lignocellulosic biomass is the world's most abundant renewable organic carbon-based resource, and it is considered the most promising alternative to fossil resources (Luterbacher *et al.* 2014). Lignocellulosic biomass consists mainly of cellulose, hemicellulose, and lignin. These three components are closely coupled with physical forces and chemical bonds to form a 3D cross-linked structure, making it challenging to separate the three components effectively (Tocco *et al.* 2021).

The authors' previous study proposed an efficient biomass pretreatment method named phosphoric acid-hydrogen peroxide (PHP) pretreatment (Wang *et al.* 2014). The systematic research established that PHP pretreatment could treat various softwoods, hardwoods, straws, and herbs with mild conditions and low grinding requirements to achieve a sufficient lignocellulose deconstruction. Under typical pretreatment conditions, the final recovered cellulose-enriched fraction (CEF) achieved a cellulose recovery of 92%, while 83.7% of the lignin and 100% of the hemicellulose were removed. After pretreatment, 29.1 to 32.6 g glucose can be harvested from 100 g wheat straw *via* enzyme hydrolysis (Wang *et al.* 2016). Through the simultaneous saccharification and fermentation (SSF) process at 15.3% solid loading for 120 h, it could harvest 15.5 g ethanol from 100 g wheat straw (Qiu *et al.* 2018). The PHP pretreatment can also separate the hemicellulose and lignin from biomass to prepare high-value products, such as oligosaccharides and supercapacitors (Wan *et al.* 2019; Liu *et al.* 2021). The oxidative tail gas from PHP pretreatment could be employed to achieve 68% to 98.3% methyl blue degradation (Lei *et al.* 2022). A recycle experiment confirmed that 86.0% phosphoric acid could be recovered after  $\geq 11$  runs of pretreatment (Yao *et al.* 2019).

Previous studies clarified the transformation mechanism of principal components and the formation of multiple oxidation systems within PHP pretreatment, which is of great significance for the in-depth development of the PHP method (Wang *et al.* 2018; Tian *et al.* 2021). However, component separation is vital in most application scenarios, which demonstrates the urgency to carry out optimization experiments to increase the component separation efficiency, especially the separation of cellulose. Response surface methodology (RSM) is a commonly used approach for response surface mapping to the region of interest, response optimization, and operation conditions selection (Pereira *et al.* 2021). RSM with Box-Behnken design was employed to select PHP pretreatment conditions, and significant improvement was achieved. For example, cellulose yield maximally achieved 946.2 mg/g after 72 h enzymatic hydrolysis, under the optimized conditions of 40°C, 2.0 h, and 70.2% H<sub>3</sub>PO<sub>4</sub> (Qiu *et al.* 2017). However, both lignocellulose deconstruction and lignin/hemicellulose degradation processes are complex and non-linear. They pose a challenge for prediction with RSM, especially with limited experimental groups. RSM can only achieve good prediction accuracy within a limited range of pretreatment conditions. When the range of pretreatment conditions is large, the prediction results are often unsatisfactory.

Nowadays, a powerful predictive tool named Artificial Neural Network (ANN) is employed in various research areas because of its modeling ability, even if limited experimental data are provided (Rashid *et al.* 2021). The ANN technology is enlightened by the operational mode of the human brain and nervous system, which contains numerous neurons (also known as processing elements or perceptron) in multiple layers. A neuron may connect to all or a subset of the neurons in the subsequent layer, with these connections simulating the brain's synaptic connections (Walczak and Cerpa 2001). For this reason, ANN can learn from complex, linear, and non-linear systems without any prior fitting function specified (Rashid *et al.* 2021). A recent study collected a total of 482 samples, and it evaluated the primary pyrolysis products of lignocellulosic biomass via ANN modeling; this approach successfully achieved the best possible results over different reactor systems, conditions, and biomass for the solid, liquid, and gaseous pyrolysis product yields (Tsekos *et al.* 2021). A prediction of phenolic compounds/glucose content from dilute inorganic acid pretreatment of biomass was proposed, which suggests that the ANN model could predict the pretreatment efficiency with limited conditions and groups of pretreatments

(Luo *et al.* 2021). Moreover, the applications of the ANN model in biomass components estimation (Kartal and Özveren 2021), kinetic parameters prediction of biomass oxidation (Sunphorka *et al.* 2017), and pretreatment for lignocellulose degradation (Bhange *et al.* 2017) were studied. And these applications fully demonstrate the potential of the ANN model in biomass valorization. This study aimed to develop an ANN model to estimate the efficiency of PHP pretreatment. Four significant factors, time, temperature, H<sub>3</sub>PO<sub>4</sub> concentration, and H<sub>2</sub>O<sub>2</sub> concentration, were used as input variables. The output variables included cellulose content and recovery, hemicellulose removal, and lignin removal. The relative importance of these pretreatment conditions was evaluated by analyzing the neural net weights in the developed ANN model.

## EXPERIMENTAL

### Materials and Methods

#### Materials

Wheat straw (WS) was collected from a farm at Sichuan Agricultural University (Chengdu, China). The WS was air-dried and milled to pass through a 40-mesh sieve (< 0.45 mm) before undergoing PHP pretreatment. All reagents used were of analytical grade and were obtained from Sigma-Aldrich (Shanghai, China) unless stated otherwise.

#### Methods

##### PHP pretreatment

To carry out the PHP pretreatment, a PHP solution was prepared by diluting 85% H<sub>3</sub>PO<sub>4</sub> with 30% H<sub>2</sub>O<sub>2</sub>. Wheat straw was then added to the solution at a solid/liquid ratio of 1:10 (w/w) in a 250 mL screw-cap bottle and mixed thoroughly. The mixture was shaken at the designed pretreatment temperature and reaction time under a rotation speed of 160 r/min. After reaching the set time, 1.0 L of deionized water was added to stop the pretreatment. The treated WS was then filtered and washed to a neutral pH to obtain the cellulose-enriched fraction (CEF), which was frozen at -20°C until further use. It should be noted that the data for modeling (training and validation of the ANN model) were derived from previous experimental results, while the data for model testing were newly collected in this study under different pretreatment conditions from the previous ones. Furthermore, this study will keep collecting more data on PHP pretreatment and enhancing the ANN model proposed herein to obtain better prediction outcomes. The PHP pretreatment conditions is listed in Table 1.

**Table 1.** PHP Pretreatment Conditions

Purpose	t (h)	T (°C)	C <sub>p</sub> (%)	C <sub>h</sub> (%)
ANN training and validation	2 – 4	30 – 50	65 – 85	0 – 7.06
ANN testing	2	45	65 – 85	0 – 7.06

Note: The raw data obtained for ANN training, validation, and testing are listed in Tables S1 and S2, respectively.

##### Analytical methods

The main components of WS material and CEF, including cellulose, hemicellulose, and lignin were analyzed according to the NREL method (National Renewable Energy Laboratory of the US) (Sluiter *et al.* 2010). The hydrolysate sugars were separated using a

Shodex SH1011 column at 60 °C with a 0.05 mol/L H<sub>2</sub>SO<sub>4</sub> mobile phase at a flow rate of 0.8 mL/min. The separated sugars were then quantified using an Agilent 1260 Infinity II HPLC system with a G7162A differential refractive index detector (Agilent Technologies, Santa Clara, CA, USA). The solid recovery (SR) of WS after pretreatment was calculated using the Eq. 1,

$$SR (\%) = m_1/m_0 \times 100\% \quad (1)$$

where  $m_0$  represents the dry weight of WS (in this case,  $m_0 = 5.00$  g), and  $m_1$  represents the dry weight of CEF (g).

The cellulose recovery ( $R_y$ ) and removal of hemicellulose or lignin ( $R_l$ ) after PHP pretreatment were calculated using the following equations:

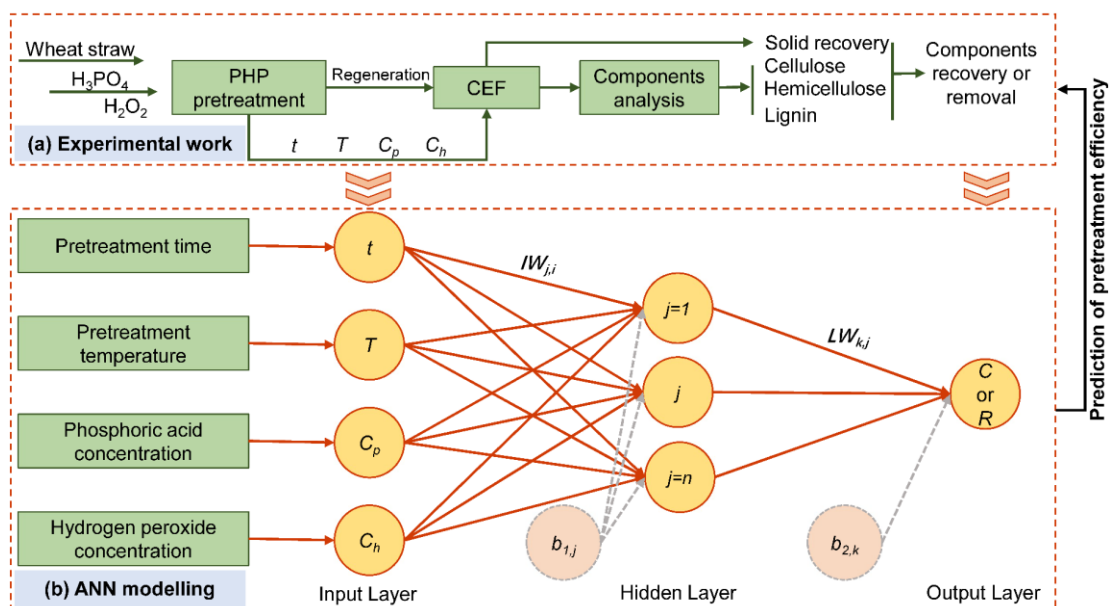
$$R_y (\%) = SR \times (C_1 / C_0) \times 100\% \quad (2)$$

$$R_l (\%) = 100\% - R_y \quad (3)$$

where  $C_0$  and  $C_1$  represent the contents of the related components in WS and CEF, respectively.

### ANN modeling

Figure 1a shows the flow diagram of PHP pretreatment. The primary parameters that affect the efficiency of PHP pretreatment are time ( $t$ ), temperature ( $T$ ), concentration of phosphoric acid ( $C_p$ ), and concentration of hydrogen peroxide ( $C_h$ ). After pretreatment, the main components of the obtained CEF, including solid recovery, the content of cellulose/hemicellulose/lignin as well as their corresponding recovery and removal percentages, were analyzed simultaneously (see Table S1 for the results).



**Fig. 1.** The flowchart of PHP pretreatment and the development of the ANN model: (a) Experimental work illustrates the procedure and data acquisition for PHP pretreatment; (b) The topology of a three-layer ANN model for predicting cellulose content and recovery of CEF.  $b_{1,j}$  represents the bias of inputs;  $b_{2,k}$  represents the bias of outputs;  $IW_{j,i}$  represents weights from inputs to hidden layers;  $LW_{k,j}$  represents weights from neurons in hidden layers to output layers.

The ANN model was developed using the Neural Network Fitting app (version 1.33) based on the neural net library in R software and implemented in OriginPro 2022 (OriginLab Corp., Northampton, MA, USA).

As shown in Fig. 1b, the ANN model proposed in this study is a multilayer perceptron (MLP) network. The MLP is a class of feedforward artificial neural network, that consists of an input layer with four input variables, a hidden layer, and an output layer with one output variable. The neurons in each layer are fully connected to the neurons in the following layer.

The number of neurons ( $n$ ) in the hidden layer was determined using an empirical equation (Yang *et al.* 2020) as follows,

$$n \leq \sqrt{i + k} + \alpha \quad (4)$$

where  $n$  represents the number of neurons in the hidden layer,  $i$  represents the number of input variables,  $k$  represents the number of output variables, and  $\alpha$  is a constant ranging from 1 to 10. The accuracy of the ANN modeling and predictions was assessed using the Root Means Square Error (RMSE), which was calculated using the following equation,

$$RMSE = \sqrt{\frac{1}{m} \sum_{h=1}^m (y_{pre}^{(h)} - y_{exp}^{(h)})^2} \quad (5)$$

where  $y_{pre}^{(h)}$  represents the predicted value obtained from the ANN model and  $y_{exp}^{(h)}$  represents the experimental value;  $m$  is the number of samples used for ANN modeling. The relative importance of the four input variables on each output variable was evaluated using the Garson's equation,

$$I_i = \frac{\sum_{j=1}^n \left( \frac{|LW_{j,i}|}{\sum_{i=1}^4 |LW_{j,i}|} \times |LW_{k,j}| \right)}{\sum_{i=1}^4 \left\{ \sum_{j=1}^n \left( \frac{|LW_{j,i}|}{\sum_{i=1}^4 |LW_{j,i}|} \times |LW_{k,j}| \right) \right\}} \times 100\% \quad (6)$$

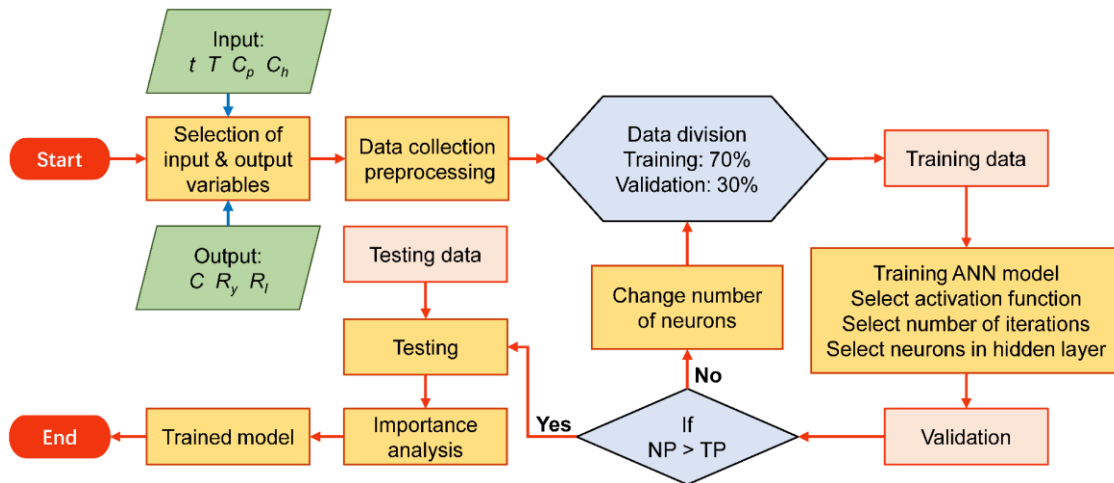
where  $I_i$  represents the relative importance of the  $i^{\text{th}}$  input variable on the output variable;  $LW_{j,i}$  represents the net weight from  $i^{\text{th}}$  input variable to  $j^{\text{th}}$  neuron in the hidden layer; and  $LW_{k,j}$  represents the net weight from the  $j^{\text{th}}$  neuron in the hidden layer to the  $k^{\text{th}}$  output variable.

**Table 2.** Selection of Modeling Parameters

Model Parameters	Specifications
Number of input variables	4 ( $t$ , $T$ , $C_p$ , $C_h$ )
Number of output variables	1 ( $C$ , $R_y$ , or $R_l$ )
Network algorithm	Resilient backpropagation with backtracking
Activation functions	Logistic, ReLU, or Tanh
Number of iterations	5 to 1000
Number of hidden layers	1
Number of neurons in hidden layers	1 to 13
Error function	RMSE
Threshold of the error function	0.02

Table 2 lists the key parameters of an ANN model. The activation functions, number of iterations, and number of neurons in hidden layers still need to be determined within a given range. As Fig. 2 illustrates, the entire dataset derived from previous

experimental results (see Table S1) was split into a 7:3 ratio for ANN training and validation. The ANN model was constructed by selecting an activation function, choosing the number of iterations, and selecting the number of neurons in the hidden layer. The RMSE served as the main indicator for comparing NP and TP. The trained ANN model was then tested with a new dataset (see Table S2) that had different pretreatment conditions from the previous ones to assess its accuracy in real applications. The *RI* values were also calculated using the final proposed ANN model to better understand each parameter's role in PHP pretreatment.



**Fig. 2.** Flowchart of the ANN modeling; NP, network performance; TP, target network performance; C, component content;  $R_y$ , component recovery rate;  $R_i$ , component removal rate

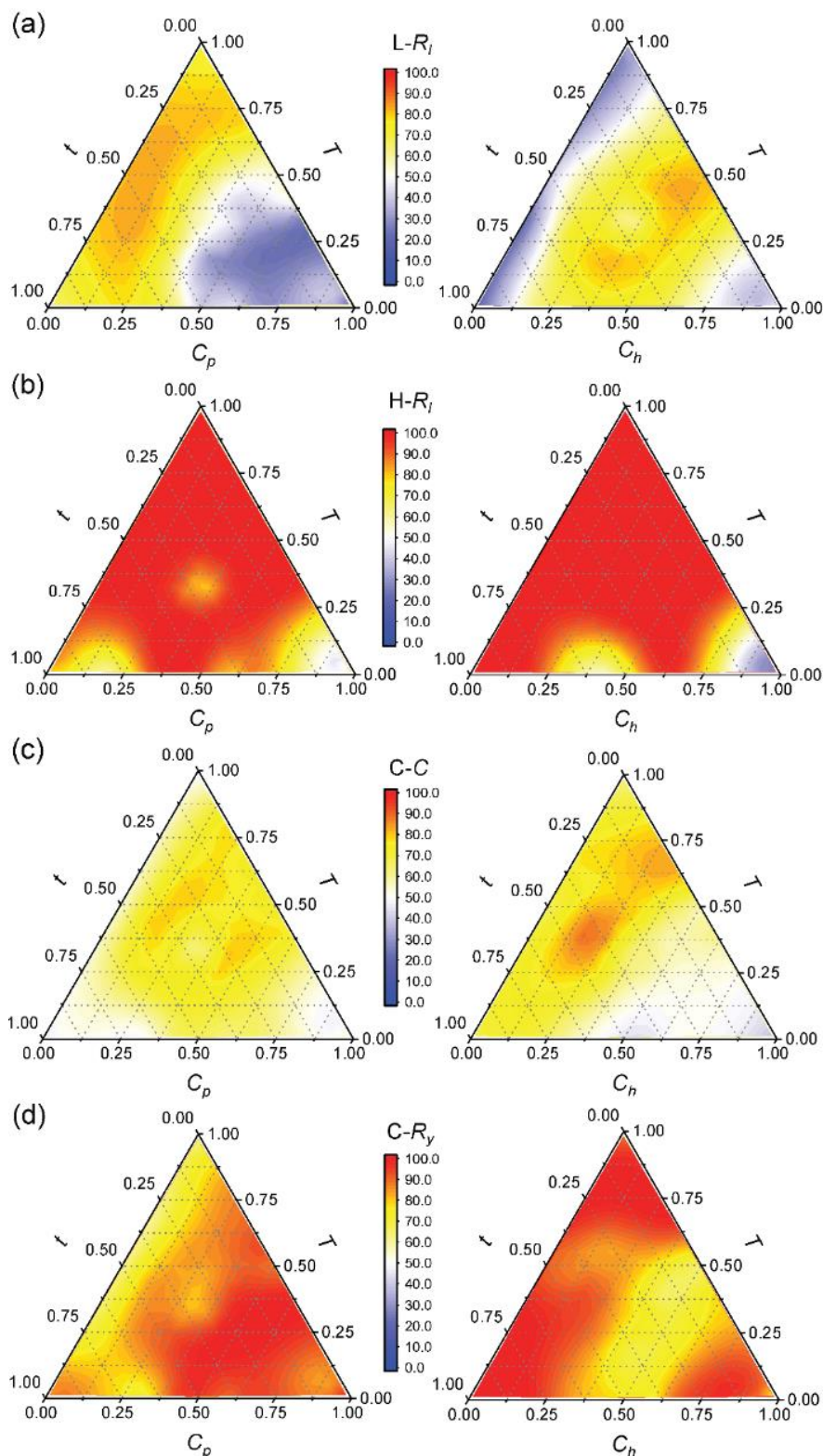
### Statistical analysis

The raw data was digitized and preprocessed in Excel (Microsoft Corp., Redmond, WA, USA), where formula calculations, such as component recovery/removal and RMSE values, were performed. One-way analysis of variance (ANOVA) was carried out using OriginPro (OriginLab Corp., Northampton, MA, USA) software. Data are displayed as means  $\pm$  standard deviation (SD) and differences among means were compared using Fisher's Least Significant Difference (LSD) method at the significance level of "ns"  $P > 0.05$ , \*  $P < 0.05$ , \*\*  $P < 0.01$ , and \*\*\*  $P < 0.001$ .

## RESULTS AND DISCUSSION

### Pretreatment of WS under Various Conditions

PHP pretreatment was performed for WS based on a single-factor experiment design within a predefined parameter range. The contents of cellulose, hemicellulose, and lignin were analyzed in the recovered CEF according to well-established procedures (Wang *et al.* 2016). The removal of lignin and hemicellulose is a critical evaluation index of pretreatment efficiency because they form a protective physical barrier to general valorization applications (Ohgren *et al.* 2007). Figure 3a shows how  $L-R_1$  varied with the changes of  $T$  and  $C_p$  ( $C_h$ ). The authors observed that high  $T$  and  $C_h$  led to strong delignification.



**Fig. 3.** Effect of PHP pretreatment conditions on CEF components: (a) L-R<sub>l</sub>, lignin removal; (b) H-R<sub>l</sub>, hemicellulose removal; (c) C-C, cellulose content; (d) C-R<sub>y</sub>, cellulose recovery. The parameters of each PHP condition are normalized in [0, 1] to draw a ternary contour plot. The color mapping refers to an actual percentage of component content, recovery, or removal. Each evaluation index is exhibited by two ternary contour plots, with the abscissa being either H<sub>3</sub>PO<sub>4</sub> concentration or H<sub>2</sub>O<sub>2</sub> concentration.

However, when  $C_h$  was close to 100%, lignin removal decreased significantly. Moreover,  $L-R_l$  did not change much as  $t$  increased. These results indicate that  $T$ ,  $C_p$ , and  $C_h$  may dominate lignin removal in PHP pretreatment. Figure 3b shows the distribution of  $H-R_l$  under various PHP conditions. Hemicellulose removal sharply increased to 100% when  $t$  and  $T$  increased, suggesting its sensitivity to  $t$  and  $T$  in PHP pretreatment.

The purpose of pretreatment is to separate cellulose for more straightforward utilization. Thus, cellulose purity and yield are the primary evaluation indicators that should be considered carefully (Tang *et al.* 2019). In this study, the authors determined two related indicators:  $C-C$  and  $C-R_y$ , as shown in Figs. 3c and 3d, respectively. As depicted in Fig. 3c,  $C-C$  was relatively low under low pretreatment intensity (short  $t$  or low  $T$ ) or excessive  $C_p/C_h$ . From the comprehensive analysis of the  $L-R_l$  and  $H-R_l$ , the relative content of cellulose was low because of the presence of lignin and hemicellulose under low pretreatment intensity. The  $C-C$  increases when pretreated with severe conditions (high  $T$  or long  $t$ ), and  $C-R_y$  decreased significantly because of oxidative degradation or acid hydrolysis (Fig. 3d). Therefore, estimating PHP pretreatment efficiency is a complicated process that involves balancing barrier component removal and cellulose degradation. This process makes all essential PHP conditions vital to  $C-C$  and  $C-R_y$ .

### Training/validation for the ANN model

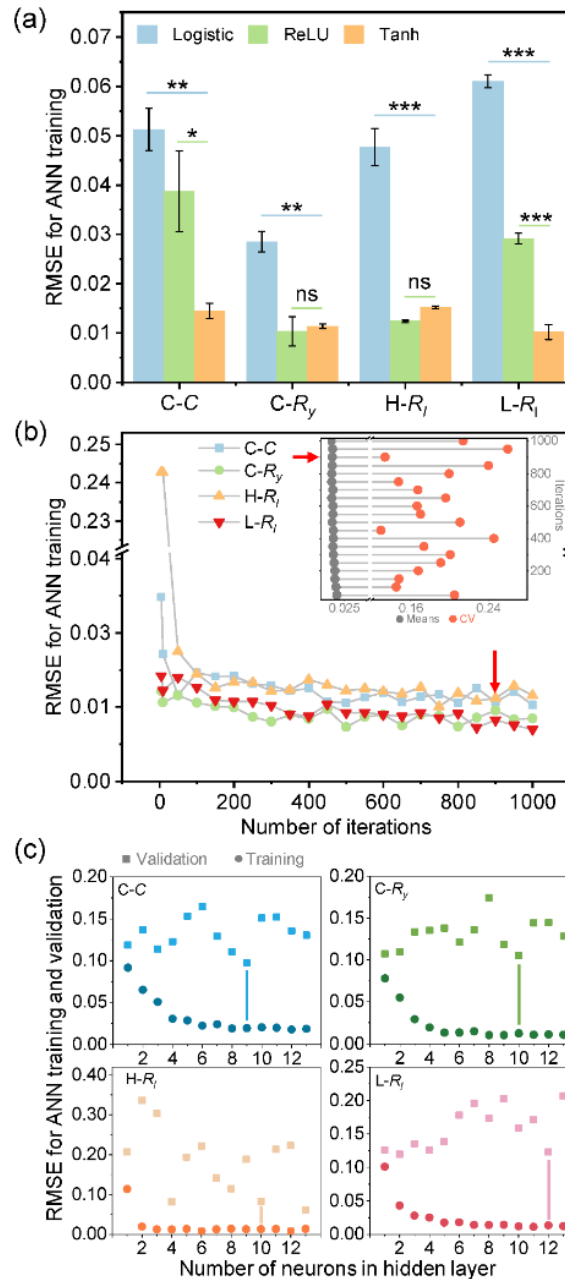
Pretreatment methods that target specific components can help reduce energy and chemical consumption and achieve multi-stage utilization of biomass (Wagle *et al.* 2022). However, finding the optimal pretreatment conditions for different types of biomasses is a complex task that requires accurate and reliable evaluation models. In this study, the authors proposed a novel model based on ANN technology to predict pretreatment efficiency for WS. The authors used four pretreatment conditions ( $t$ ,  $T$ ,  $C_p$ , and  $C_h$ ) as input variables and four obtained results ( $C-C$ ,  $C-R_y$ ,  $H-R_l$ , and  $L-R_l$ ) as output variables for our ANN model.

The ANN model consists of three parts: interconnections, activation functions, and learning rules (Sadiq *et al.* 2019). The authors used a multilayer feed-forward network with resilient backpropagation with backtracking as the learning rule (see Fig. 1 and Table 2). The main challenge is to select a suitable activation function. Activation functions enable deep neural networks to learn complex mappings between input and output. Without them, a neural network can only learn a linear relation (Goyal *et al.* 2020).

The authors compared different activation functions for PHP pretreatment: rectified linear units (ReLU) and sigmoid functions such as logistic and tangent hyperbolic (tanh). Figure 4a shows the RMSE value for ANN training with each activation function. Tanh performed significantly better than logistic for  $C-C$ ,  $C-R_y$ ,  $H-R_l$ , and  $L-R_l$  output variables ( $P < 0.01$  or  $P < 0.001$ ). Tanh also outperformed ReLU for  $C-C$  ( $P < 0.05$ ) and  $L-R_l$  ( $P < 0.001$ ) output variables, but not for  $C-R_l$  and  $H-R_l$  output variables ( $P > 0.05$ ). Therefore, the authors chose tanh as the ideal activation function for ANN modeling of PHP pretreatment.

Figure 4b shows the RMSE values under the different number of iterations. As the number of iterations increases ( $\geq 550$  iterations), the RMSE decreases to a relatively stable state. The coefficient of variation (CV) is employed to estimate the differences between various output variables (refer to the subplot). The number of iterations needed to find an optimal solution for a given accuracy largely determines the overall computational efforts and the performance of an algorithm.



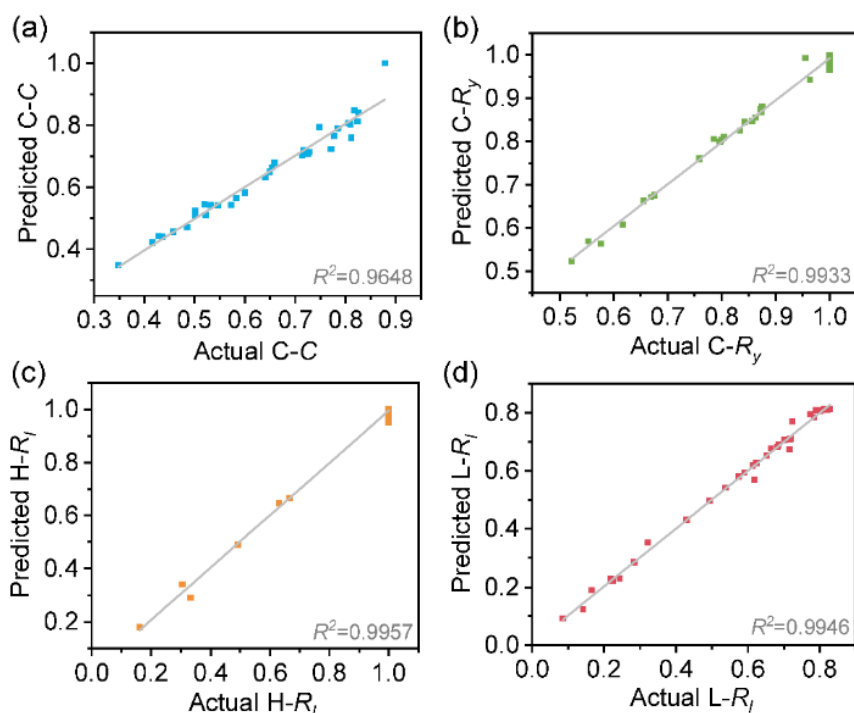


**Fig. 4.** RMSE values of C-C, C-R<sub>y</sub>, H-R<sub>l</sub>, and L-R<sub>l</sub> for the ANN training and validation. The RMSE values of three activation functions (a), different numbers of iterations (b), and various numbers of neurons in the hidden layer (c)

A better ANN model should use less computation and fewer iterations (She 2014). Therefore, in this study, 900 iterations were found to be suitable for the ANN model considering its accuracy and computational amount.

Studies have shown that too many neurons in the hidden layer increase the training time and make it difficult to achieve the desired effect, even if the training data contains enough information (Panchal *et al.* 2011). Therefore, choosing a suitable number of neurons in the hidden layer that minimizes the error rate and maximizes the generalization ability is crucial. Figure 4c exhibits the RMSE of training/validation for four ANN models with different numbers of neurons under 900 iterations. The range of neuron numbers (1 to

13) was confirmed by an empirical equation (Eq. 4). As the iteration times increased, the RMSE of ANN training gradually decreased until reaching a stable state (the demarcation is 8th, 8th, 6th, and 11th for C-C, C-R<sub>y</sub>, H-R<sub>l</sub>, and L-R<sub>l</sub>, respectively). However, its RMSE for ANN validation changes with different neuron numbers. According to the previous study, the general idea is that fewer neurons will underfit, whereas too many neurons will overfit an ANN model; this is true when the other factors such as the complexity of the problem, the number of layers, the activation functions, and the number of neurons in each layer are fixed (Adil *et al.* 2022). To avoid these circumstances, the authors employed a facility principle of “the minimum RMSE value with minimum neurons” to determine the number of neurons for each output variable. The final option is 9 neurons for C-C, 10 neurons for C-R<sub>y</sub> and H-R<sub>l</sub>, and 12 neurons for L-R<sub>l</sub>. Figure 5 compares experimental results with predicted results obtained from the optimized parameters: all data are well fitted with an R<sup>2</sup> value ranging from 0.9648 to 0.9957, which indicates excellent prediction accuracy.



**Fig. 5.** Comparison of the experimental results and predicted results derived from the ANN model for (a) C-C, (b) C-R<sub>y</sub>, (c) H-R<sub>l</sub>, and (d) L-R<sub>l</sub> in the training periods

### Prediction and Relative Importance of Input Variables

The optimal ANN structure was determined through numerical experiments using training and validation datasets (see Table S1 for complete datasets). A separate group of the experiment (see Table 1 for conditions) was conducted to verify the effectiveness of the trained ANN model. For the results of testing, see Table S2. Figure 6a displays the fitting relationships between predicted and experimental values for C-C, C-R<sub>y</sub>, H-R<sub>l</sub>, and L-R<sub>l</sub>. The slope (*S*) and coefficient of determination (R<sup>2</sup>) of the fit curve were used to evaluate the accuracy of the trained ANN model (Luo *et al.* 2021). Values near 1.00 for both *S* and R<sup>2</sup> indicate high accuracy in predicting a specific variable. As shown in Fig. 6a, C-C, H-R<sub>l</sub>, and L-R<sub>l</sub> achieved excellent modeling precision with *S* values ranging from 0.96 to 1.07 and R<sup>2</sup> values from 0.9917 to 0.9989. C-C had an *S* value of 0.63 and an R<sup>2</sup> value

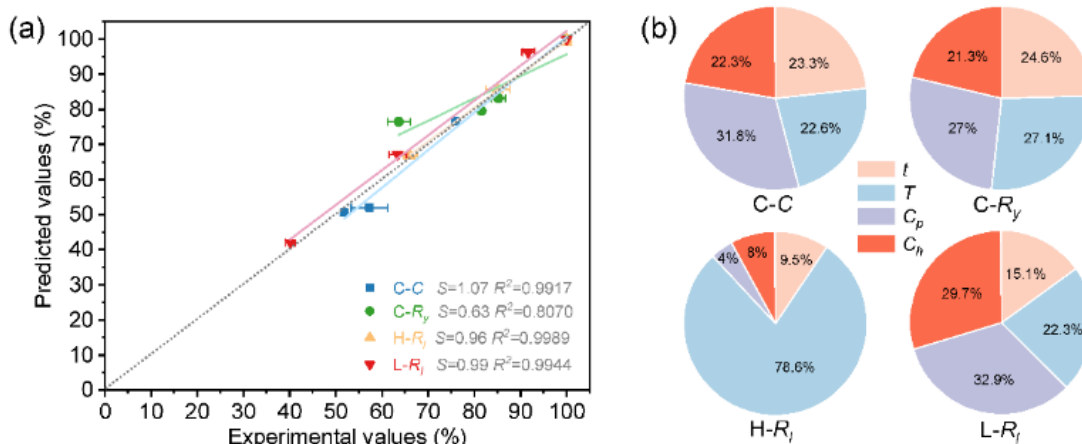


	10	0.4681	-	0.3928	-0.2164	0.6000	-	
			18.0798				0.2335	
<b>Lignin removal (L-R)</b>	1	-0.0528	-1.0832	-4.2216	1.3689	-	-	-
						1.1109	1.1378	0.2323
	2	8.5965	-6.8695	-0.0166	-	0.6172	0.2271	
					16.5425			
	3	-0.2676	-0.8281	1.3845	0.4985	-	-	
						0.8440	2.1040	
	4	-0.7472	-0.1430	-0.1792	0.5669	0.9663	1.1065	
	5	-2.6061	-5.2002	-0.9199	3.2301	-	-	
						0.0599	0.5832	
	6	-0.0201	0.1240	-0.5098	0.1887	0.9007	1.4053	
	7	-0.5294	4.6038	0.2552	-	0.3497	-	
					14.3484		0.4417	
8	-1.1109	25.3572	1.3814	15.8228	-	-		
					0.2981	0.3107		
9	0.2910	0.7686	0.6973	-0.6124	0.5303	-		
						2.5000		
10	-0.4854	0.1353	0.8392	2.0677	-	-		
					1.2080	1.2096		
11	-9.3648	-1.3826	-1.1954	0.5778	0.2294	-		
						0.2585		
12	-3.4432	-1.1727	1.1569	-2.2897	0.3578	0.3301		

Weights and biases are learnable parameters in an ANN model. Weights determine the influence of inputs on the output, while biases adjust the output and weighted sum of inputs to a neuron. Table 3 shows the weights ( $IW_{j,i}$  and  $LW_{k,j}$ ) and biases ( $b_{1,j}$  and  $b_{2,k}$ ) from 4 input variables ( $t$ ,  $T$ ,  $C_p$ , and  $C_h$ ) to 1 output variable ( $C-C$ ,  $C-R_y$ ,  $H-R_t$ , or  $L-R_t$ ). Relative importance ( $RI$ ) is a key indicator that describes the influence of each input variable on the output variable. It is calculated using the Garson equation, as shown in Fig. 6b.

The pie chart shows that the  $RI$  of  $C_p$  on  $C-C$  was 31.8%, which was higher than other variables. However, the remaining three variables had almost equal  $RI$  and also played a significant role in  $C-C$ . For  $C-R_y$ ,  $T$  (27.1%) and  $C_p$  (27.0%) had higher  $RI$  than  $t$  (24.6%) and  $C_h$  (21.3%). In  $H-R_t$ ,  $T$  had an  $RI$  of 78.6%, which was significantly higher than other variables. For  $L-R_t$ ,  $C_p$  and  $C_h$  had more impact than  $T$  and  $t$ . These results indicate that the concentration of  $H_3PO_4$  and  $H_2O_2$  (with total  $RI$  for each output variable ranging from 48.3% to 62.6%) had a significant effect on pretreatment efficiency. The different  $RI$  values for various output variables suggest a complex synergistic effect between  $H_3PO_4$  and  $H_2O_2$  on PHP pretreatment. This implies that there was still room for improvement in adjusting the concentration ratio of  $H_3PO_4/H_2O_2$  by diluting 85% phosphoric acid with 30% hydrogen peroxide.

Hemicellulose in WS was found to be more sensitive to changes in pretreatment temperature than cellulose and lignin. According to Fig. 6b, pretreatment time (with  $RI$  ranging from 9.5% to 24.6% for each output variable) did not appear to be the main factor affecting PHP pretreatment efficiency. Therefore, future studies should focus on understanding the composition-efficacy relationship between  $H_3PO_4$  and  $H_2O_2$  and refining pretreatment conditions based on current research.



**Fig. 6.** Testing the trained ANN model (a) and  $R/I$  values of PHP pretreatment conditions on each output variable (b)

The ANN modeling outperformed the RSM in dealing with the data with a large amount or/and many parameters. Typically, when performing the RSM with Box-Behnken design, the program designs the pretreatment conditions according to the boundary constraints, and the experiment has to adhere to the experimental design table strictly. Furthermore, adding an extra parameter or refining the experimental condition will cause a drastic increase in the number of experimental runs. When there is a nonlinear response within the condition boundary, such as a chain reaction triggered by a certain condition that considerably improves the decomposition extent, the RSM fails to provide valid results. On the contrary, the ANN modeling has minimal data requirements, and the abundant data accumulated in the pre-experiments can also be utilized for model training, which is a very convenient method for biomass pretreatment. Thus, the present results suggest that ANN combined with the AI technology can play a remarkable role in the traditional biomass pretreatment field.

## CONCLUSIONS

1. This paper explored the feasibility of using ANNs to predict PHP pretreatment efficiency. It was experimentally verified that the tanh activation function is suitable for modeling in this study. Each trained ANN model had 4 key conditions as input variables, one hidden layer, and one selected output variable. The specific network trained for cellulose content, cellulose recovery, hemicellulose removal, and lignin removal had 9, 10, 10, and 12 neurons, respectively.
2. After 900 iterations, their modeling accuracy for validation ( $R^2$ ) reached 0.9648 to 0.9957. The  $R^2$  of testing datasets ranged from 0.8070 to 0.9989, indicating excellent predictive efficiency of the proposed ANN models.
3. The relative importance of four pretreatment conditions to pretreatment efficiency was also investigated to provide insights for fine-tuning PHP pretreatment. In summary, considering the fitting accuracy between predicted and experimental results, ANN-based models were judged to be beneficial for predicting PHP pretreatment efficiency.

## ACKNOWLEDGMENTS

The authors are grateful to the China Scholarship Council (CSC, 201906910047), the University of Calgary, the National Natural Science Foundation of China (21978183), and the Science and Technology Department of Sichuan Province (2022YFH0065, 2022YFN0027, 2022NSFSC0394, and 2021ZYD0099) for supporting this study. The authors thank the Biorefining and Photo-Bioprocessing Research Laboratory for providing the necessary reagents, instruments, and technical assistance.

## REFERENCES CITED

- Adil, M., Ullah, R., Noor, S., and Gohar, N. (2022). "Effect of number of neurons and layers in an artificial neural network for generalized concrete mix design," *Neural Comput & Applic* 34, 8355-8363. DOI: 10.1007/s00521-020-05305-8
- Bhange, V. P., Bhivgade, U. V., and Vaidya, A. N. (2017). "Artificial neural network modeling in pretreatment of garden biomass for lignocellulose degradation," *Waste Biomass Valorization* 10(5), 1571-1583. DOI: 10.1007/s12649-017-0163-z
- Goyal, M., Goyal, R., Venkatappa Reddy, P., and Lall, B. (2020). "Activation functions, in: *Deep Learning: Algorithms and Applications*, W. Pedrycz, S. M. Chen (eds.), Springer, Cham, Switzerland, pp. 1-30. DOI: 10.1007/978-3-030-31760-7\_1
- Hosseini Koupaie, E., Dahadha, S., Bazyar Lakeh, A. A., Azizi, A., and Elbeshbishy, E. (2019). "Enzymatic pretreatment of lignocellulosic biomass for enhanced biomethane production-a review," *J. Environ. Manag.* 233, 774-784. DOI: 10.1016/j.jenvman.2018.09.106
- Kartal, F., and Özveren, U. (2021). "An improved machine learning approach to estimate hemicellulose, cellulose, and lignin in biomass," *Carbohydr. Polym. Technol. Appl.* 2(25), article ID 100148. DOI: 10.1016/j.carpta.2021.100148
- Lei, M., Shen, F., Hu, J. G., Zhao, L., Huang, M., Zou, J. M., Tian, D., Yang, G., Zeng, Y. M., and Deng, S. H. (2022). "A novel way to facilely degrade organic pollutants with the tail-gas derived from PHP (phosphoric acid plus hydrogen peroxide) pretreatment of lignocellulose," *J. Hazard. Mater.* 424(Part B), article ID 127517. DOI: 10.1016/j.jhazmat.2021.127517
- Liu, Z. L., Wan, X., Wang, Q., Tian, D., Hu, J. G., Huang, M., Shen, F., and Zeng, Y. M. (2021). "Performances of a multi-product strategy for bioethanol, lignin, and ultra-high surface area carbon from lignocellulose by PHP (phosphoric acid plus hydrogen peroxide) pretreatment platform," *Renew. Sust. Energ. Rev.* 150, article ID 111503. DOI: 10.1016/j.rser.2021.111503
- Luo, H. Z., Gao, L., Liu, Z., Shi, Y. J., Xie, F., Bilal, M., Yang, R. L., and Taherzadeh, M. J. (2021). "Prediction of phenolic compounds and glucose content from dilute inorganic acid pretreatment of lignocellulosic biomass using artificial neural network modeling," *Bioresour. Bioprocess.* 8(1), article number 134. DOI: 10.1186/s40643-021-00488-x
- Luterbacher, J. S., Martin Alonso, D., and Dumesic, J. A. (2014). "Targeted chemical upgrading of lignocellulosic biomass to platform molecules," *Green Chem.* 16(12), 4816-4838. DOI: 10.1039/c4gc01160k

- Ohgren, K., Bura, R., Saddler, J., and Zacchi, G. (2007). "Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover," *Bioresource Technol.* 98(13), 2503-2510. DOI: 10.1016/j.biortech.2006.09.003
- Panchal, G., Ganatra, A., Kosta, Y. P., and Panchal, D. (2011). "Behaviour analysis of multilayer perceptron with multiple hidden neurons and hidden layers," *Int. J. Comput Theor. Eng.* 3(2), 332-337. DOI: 10.7763/ijcte.2011.V3.328
- Pereira, L. M. S., Milan, T. M., and Tapia-Blácido, D. R. (2021). "Using response surface methodology (RSM) to optimize 2G bioethanol production: A review," *Biomass Bioenergy* 151, article ID 106166. DOI: 10.1016/j.biombioe.2021.106166
- Qiu, J. W., Tian, D., Shen, F., Hu, J. G., Zeng, Y. M., Yang, G., Zhang, Y. Z., Deng, S. H., and Zhang, J. (2018). "Bioethanol production from wheat straw by phosphoric acid plus hydrogen peroxide (PHP) pretreatment via simultaneous saccharification and fermentation (SSF) at high solid loadings," *Bioresource Technol.* 268, 355-362. DOI: 10.1016/j.biortech.2018.08.009
- Qiu, J. W., Wang, Q., Shen, F., Yang, G., Zhang, Y. Z., Deng, S. H., Zhang, J., Zeng, Y. M., and Song, C. (2017). "Optimizing phosphoric acid plus hydrogen peroxide (PHP) pretreatment on wheat straw by response surface method for enzymatic saccharification," *Appl. Biochem. Biotechnol.* 181(3), 1123-1139. DOI: 10.1007/s12010-016-2273-7
- Rashid, T., Ali Ammar Taqvi, S., Sher, F., Rubab, S., Thanabalan, M., Bilal, M., and Ul Islam, B. (2021). "Enhanced lignin extraction and optimisation from oil palm biomass using neural network modelling," *Fuel* 293, article ID 120485. DOI: 10.1016/j.fuel.2021.120485
- Sadiq, R., Rodriguez, M. J., and Mian, H. R. (2019). "Empirical models to predict disinfection by-products (DBPs) in drinking water: An updated review," in: *Encyclopedia of Environmental Health*, J. O. Nriagu (Ed.), Elsevier, Berkeley, CA, USA. DOI: 10.1016/b978-0-12-409548-9.11193-5
- She, Y. X. (2014). *Nature-Inspired Optimization Algorithms*, 2nd ed., Elsevier, Amsterdam, Netherlands. DOI: 10.1016/c2019-0-03762-4
- Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter, J., Templeton, D., and Crocker, D. (2010). "Determination of structural carbohydrates and lignin in biomass," (<https://www.nrel.gov/docs/gen/fy13/42618.pdf>), Accessed 12 June 2023.
- Sunphorka, S., Chalermainsuwan, B., and Piumsomboon, P. (2017). "Application of artificial neural network for kinetic parameters prediction of biomass oxidation from biomass properties," *J. Energy Inst.* 90(1), 51-61. DOI: 10.1016/j.joei.2015.10.007
- Tang, S., Dong, Q., Fang, Z., and Miao, Z. D. (2019). "Complete recovery of cellulose from rice straw pretreated with ethylene glycol and aluminum chloride for enzymatic hydrolysis," *Bioresource Technol.* 284, 98-104. DOI: 10.1016/j.biortech.2019.03.100
- Tian, D., Chen, Y. Y., Shen, F., Luo, M. Y., Huang, M., Hu, J. G., Zhang, Y. Z., Deng, S. H., and Zhao, L. (2021). "Self-generated peroxyacetic acid in phosphoric acid plus hydrogen peroxide pretreatment mediated lignocellulose deconstruction and delignification," *Biotechnol. Biofuels* 14(1), article ID 224. DOI: 10.1186/s13068-021-02075-w
- Tocco, D., Carucci, C., Monduzzi, M., Salis, A., and Sanjust, E. (2021). "Recent developments in the delignification and exploitation of grass lignocellulosic biomass," *ACS Sustain. Chem. Eng.* 9(6), 2412-2432. DOI: 10.1021/acssuschemeng.0c07266

- Tsekos, C., Tandurella, S., and De Jong, W. (2021). "Estimation of lignocellulosic biomass pyrolysis product yields using artificial neural networks," *J. Anal. Appl. Pyrolysis* 157(9), article ID 105180. DOI: 10.1016/j.jaap.2021.105180
- Wagle, A., Angove, M. J., Mahara, A., Wagle, A., Mainali, B., Martins, M., Goldbeck, R., and Raj Paudel, S. (2022). "Multi-stage pre-treatment of lignocellulosic biomass for multi-product biorefinery: A review," *Sustain. Energy Tech. Assmt.* 49(20), article ID 101702. DOI: 10.1016/j.seta.2021.101702
- Walczak, S., and Cerpa, N. (2001). "Artificial neural networks," in: *Encyclopedia of Physical Science and Technology*, R. A. Meyers (Ed.), Elsevier, Berkeley, CA, USA. DOI: 10.1016/b0-12-227410-5/00837-1
- Wan, X., Yao, F. P., Tian, D., Shen, F., Hu, J. G., Zeng, Y. M., Yang, G., Zhang, Y. Z., and Deng, S. H. (2019). "Pretreatment of wheat straw with phosphoric acid and hydrogen peroxide to simultaneously facilitate cellulose digestibility and modify lignin as adsorbents," *Biomolecules* 9(12), article 844. DOI: 10.3390/biom9120844
- Wang, Q., Hu, J. G., Shen, F., Mei, Z. L., Yang, G., Zhang, Y. Z., Hu, Y. D., Zhang, J., and Deng, S. H. (2016). "Pretreating wheat straw by the concentrated phosphoric acid plus hydrogen peroxide (PHP): Investigations on pretreatment conditions and structure changes," *Bioresour. Technol.* 199, 245-257. DOI: 10.1016/j.biortech.2015.07.112
- Wang, Q., Tian, D., Hu, J. G., Shen, F., Yang, G., Zhang, Y. Z., Deng, S. H., Zhang, J., Zeng, Y. M., and Hu, Y. D. (2018). "Fates of hemicellulose, lignin and cellulose in concentrated phosphoric acid with hydrogen peroxide (PHP) pretreatment," *RSC Adv.* 8(23), 12714-12723. DOI: 10.1039/c8ra00764k
- Wang, Q., Wang, Z. H., Shen, F., Hu, J. G., Sun, F. B., Lin, L. L., Yang, G., Zhang, Y. Z., and Deng, S. H. (2014). "Pretreating lignocellulosic biomass by the concentrated phosphoric acid plus hydrogen peroxide (PHP) for enzymatic hydrolysis: Evaluating the pretreatment flexibility on feedstocks and particle sizes," *Bioresour. Technol.* 166, 420-428. DOI: 10.1016/j.biortech.2014.05.088
- Yang, J., Huang, Y., Xu, H. Y., Gu, D. Y., Xu, F., Tang, J. T., Fang, C., and Yang, Y. (2020). "Optimization of fungi co-fermentation for improving anthraquinone contents and antioxidant activity using artificial neural networks," *Food Chem.* 313, article ID 126138. DOI: 10.1016/j.foodchem.2019.126138
- Yao, F. P., Tian, D., Shen, F., Hu, J. G., Zeng, Y. M., Yang, G., Zhang, Y. Z., Deng, S. H., and Zhang, J. (2019). "Recycling solvent system in phosphoric acid plus hydrogen peroxide pretreatment towards a more sustainable lignocellulose biorefinery for bioethanol," *Bioresour. Technol.* 275, 19-26. DOI: 10.1016/j.biortech.2018.12.040

Article submitted: June 2, 2023; Peer review completed: July 8, 2023; Revised version received and accepted: October 6, 2023; Published: November 15, 2023.

DOI: 10.15376/biores.19.1.288-305



## APPENDIX

## Supplementary Material

**Table S1.** Experimental Setups of PHP Pretreatment and the Obtained Results for ANN Modeling

NO.	Pretreatment Parameters				C-C (%)	C-R <sub>y</sub> (%)	H-C (%)	H-R <sub>i</sub> (%)	L-C (%)	L-R <sub>i</sub> (%)	S-R <sub>y</sub> (%)
	t (h)	T (°C)	C <sub>p</sub> (%)	C <sub>h</sub> (%)							
1	2	30	65	7.06	34.86	67.60	19.66	30.46	20.53	42.96	68.06
2	2	30	70	5.29	42.23	75.90	20.31	33.40	17.93	53.83	63.08
3	2	30	80	1.76	45.63	100.00	18.33	16.17	19.91	28.50	88.00
4	2	30	85	0.00	66.40	100.00	0.00	100.00	26.32	8.61	85.06
5	3	30	65	7.06	54.25	96.32	0.00	100.00	14.80	62.36	62.31
6	3	30	70	5.29	44.15	66.93	0.00	100.00	12.97	71.84	53.21
7	3	30	80	1.76	56.43	80.39	14.23	63.02	17.01	65.29	49.99
8	3	30	85	0.00	70.70	100.00	0.00	100.00	28.01	21.88	68.32
9	4	30	65	7.06	43.99	79.68	10.08	66.69	12.23	68.27	63.58
10	4	30	70	5.29	51.05	80.42	17.63	49.33	9.98	77.48	55.28
11	4	30	80	1.76	71.83	100.00	0.00	100.00	14.91	68.51	51.72
12	4	30	85	0.00	67.92	100.00	0.00	100.00	26.35	22.48	72.07
13	2	40	65	7.06	54.15	75.91	0.00	100.00	14.86	70.17	49.20
14	2	40	70	5.29	70.27	87.49	0.00	100.00	12.08	78.46	43.69
15	2	40	80	1.76	84.19	100.00	0.00	100.00	17.99	66.38	45.78
16	2	40	85	0.00	63.14	100.00	0.00	100.00	25.92	14.17	81.13
17	3	40	65	7.06	52.52	65.58	0.00	100.00	11.12	80.12	43.82
18	3	40	70	5.29	76.54	87.26	0.00	100.00	10.86	82.26	40.01
19	3	40	80	1.76	100.00	100.00	0.00	100.00	17.21	71.55	40.52
20	3	40	85	0.00	71.31	100.00	0.00	100.00	31.78	24.40	58.29
21	4	40	65	7.06	54.35	61.75	0.00	100.00	12.50	79.66	39.87
22	4	40	70	5.29	75.93	87.28	0.00	100.00	10.46	82.78	40.34
23	4	40	80	1.76	80.30	95.53	0.00	100.00	16.21	72.38	41.75
24	4	40	85	0.00	71.50	100.00	0.00	100.00	23.58	16.60	86.67
25	2	50	65	7.06	47.14	52.23	0.00	100.00	11.87	81.16	38.89
26	2	50	70	5.29	80.68	85.67	0.00	100.00	13.78	79.04	37.26
27	2	50	80	1.76	78.83	100.00	0.00	100.00	22.56	57.54	46.10
28	2	50	85	0.00	64.98	79.62	0.00	100.00	28.84	49.39	43.00
29	3	50	65	7.06	54.50	57.67	0.00	100.00	12.09	81.67	37.14
30	3	50	70	5.29	81.14	84.22	0.00	100.00	12.33	81.66	36.43
31	3	50	80	1.76	72.39	78.62	0.00	100.00	26.34	59.02	38.11
32	3	50	85	0.00	79.40	100.00	0.00	100.00	33.28	32.17	49.94
33	4	50	65	7.06	51.05	55.36	0.00	100.00	11.44	82.23	38.06
34	4	50	70	5.29	84.71	86.33	0.00	100.00	12.70	81.46	35.77
35	4	50	80	1.76	79.03	83.40	0.00	100.00	25.57	61.35	37.03
36	4	50	85	0.00	58.27	79.94	0.00	100.00	19.44	61.80	48.14

**Note:** The table shows the cellulose recoveries as 100% when they are calculated higher than that.

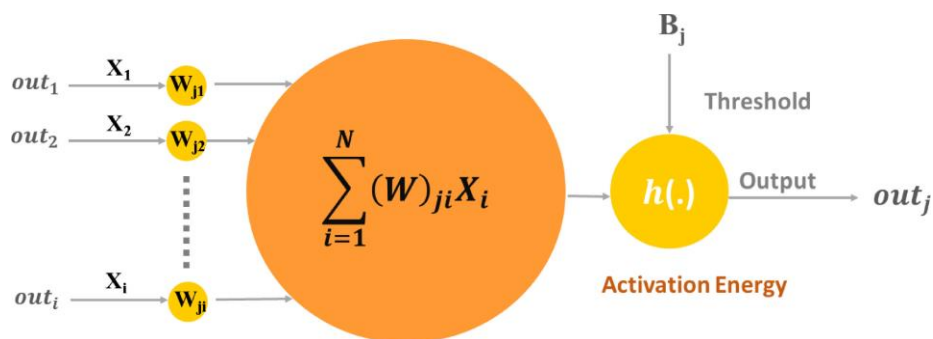
**Table S2.** Experimental Setups of PHP Pretreatment and the Obtained Results for ANN Testing

No.	Pretreatment Parameters				C-C (%)	C-R <sub>y</sub> (%)	H-C (%)	H-R <sub>l</sub> (%)	L-C (%)	L-R <sub>l</sub> (%)	S-R <sub>y</sub> (%)
	t (h)	T (°C)	C <sub>p</sub> (%)	C <sub>h</sub> (%)							
1	2	45	65	7.06	34.86	67.60	19.66	30.46	20.53	42.96	68.06
2	2	45	70	5.29	42.23	75.90	20.31	33.40	17.93	53.83	63.08
3	2	45	80	1.76	45.63	100.00	18.33	16.17	19.91	28.50	88.00
4	2	45	85	0.00	58.27	79.94	0.00	100.00	19.44	61.80	48.14

**Note:** The table shows the cellulose recoveries as 100% when they are calculated higher than that

**Table S3.** Optimized Key Parameters of the ANN Models

Output Variables	Parameters for ANN Modeling			
	Hidden Layers	Activation Functions	Iterations	Neurons
C-C	1	Tanh	900	9
C-R <sub>y</sub>	1	Tanh	900	10
H-R <sub>l</sub>	1	Tanh	900	10
L-R <sub>l</sub>	1	Tanh	900	12

**Fig. S1.** Architecture of an artificial neuron (Kartal and Özveren 2021; Creative Commons CC-BY-NC-ND)

Kartal, F., and Özveren, U. (2021). "An improved machine learning approach to estimate hemicellulose, cellulose, and lignin in biomass," *Carbohydr. Polym. Technol. Appl.* 2(25), Article ID 100148. DOI: 100148. 10.1016/j.carpta.2021.100148