

Artificial Neural Network Approach for Predicting Enzymatic Hydrolysis of Steam Exploded Pine Wood Chip in Mild Alkaline Pretreatment

Hyeon Cheol Kim , Si Young Ha , and Jae-Kyung Yang *

Lignocellulosic biomass, particularly softwoods such as pine, poses a significant challenge to enzymatic hydrolysis due to its high lignin content and complex structural rigidity. Although the application of steam explosion and alkaline pretreatment has gained widespread popularity for enhancing digestibility, the optimization of process parameters remains a formidable challenge due to the nonlinear interactions among variables. Machine learning is emerging as a promising solution to address these challenges, offering a viable alternative for predictive modeling and process control. In this study, an artificial neural network (ANN) model was developed to predict the enzymatic hydrolysis rate of steam-exploded pine wood subjected to mild alkaline (NaOH) pretreatment. The artificial neural network (ANN) was trained on experimental data encompassing three primary process variables: steam explosion time (1 to 5 min), NaOH concentration (0.5 to 2.0%), and chemical pretreatment time (12 to 24 h). The artificial neural network (ANN) model demonstrated the highest level of accuracy among the models evaluated, including random forest, support vector machine, and extreme gradient boosting. It attained a coefficient of determination (R^2) of 0.9805. In conditions that were not optimized (1% NaOH, 24-hour treatment, 5 min steam explosion, without bark), a maximum hydrolysis of 93.9% was obtained.

DOI: 10.15376/biores.20.4.8400-8419

Keywords: Artificial neural network; Alkaline pretreatment; Enzymatic hydrolysis; Pine wood; Steam explosion

Contact information: Department of Environmental Materials Science/Institute of Agriculture and Life Science, Gyeongsang National University, Jinju, 52828, Republic of Korea;

* Corresponding author: jkyang@gnu.ac.kr

INTRODUCTION

Wood cellulosic biomass feedstock is a low-cost, energy-dense, and globally abundant renewable resource (Kumar *et al.* 2020). Historically, it has been utilized as an energy source; however, the increasing dependence on fossil fuels in recent decades has led to a decline in its usage in this capacity. However, this transition has given rise to a renewed interest in its application for synthesizing high value biochemicals such as phenylpropanoids (Kawaguchi *et al.* 2016). This focus has opened new pathways for sustainable biochemical production. The primary components of wood biomass – cellulose, hemicellulose, and lignin – contain fermentable sugars (hexoses, pentoses, *etc.*) that can be converted into bioproducts through biochemical processes.

Woody biomass consists of 38 to 50% cellulose, 23 to 32% hemicellulose, and 15 to 25% lignin, although these ratios vary according to source (Vu *et al.* 2020). Enzymatic hydrolysis using cellulases has proven effective for depolymerizing cellulose into glucose, a crucial intermediate for the synthesis of value-added chemicals such as gluconic acid, glucaric acid, and levulinic acid (Zhang *et al.* 2021).

The crystalline structure of cellulose, held together by β -1,4 glycosidic bonds and extensive hydrogen bonding, poses a significant barrier to enzymatic degradation (Ding and Himmel 2006).

Nevertheless, the recalcitrance of lignocellulosic biomass remains a significant challenge. The structural rigidity of the material, attributable to the presence of hydrophobic, cross-linked lignin, has been demonstrated to impede enzyme accessibility and to reduce saccharification efficiency (Cai *et al.* 2023). This problem is especially evident in softwoods such as pine, which have a higher lignin content and more complex structures in comparison to hardwoods such as oak (Kumar *et al.* 2020). The utilization of pretreatment strategies is imperative in order to enhance the enzymatic digestibility of the lignocellulosic matrix through its fragmentation or modification.

Among pretreatment techniques, steam explosion is a widely employed method due to its ability to disrupt the lignocellulosic structure *via* high-pressure, high-temperature steam, followed by sudden decompression (Yu *et al.* 2012). This physicochemical process has been shown to increase biomass porosity and facilitate autohydrolysis of hemicellulose and partial delignification (Zabed *et al.* 2019). However, it has been demonstrated that this process can also result in the formation of fermentation inhibitors, such as furfural and 5-hydroxymethylfurfural, especially under more severe conditions. To mitigate the effects of these processes, the combination of steam explosion with chemical pretreatments is frequently employed (Nges *et al.* 2016).

Alkaline pretreatment, particularly using sodium hydroxide (NaOH), is a commonly used method to solubilize lignin and hemicellulose, reduce cellulose crystallinity, and enhance enzyme accessibility. Among alkaline reagents, NaOH is more widely used for pretreatment than potassium hydroxide (KOH) due to its lower cost and stronger alkalinity, which enables effective lignin removal even at relatively low reaction temperatures (Kim *et al.* 2016). It is cost-effective, operates under mild conditions, and can be integrated with steam explosion to synergistically improve biomass digestibility (Antonopoulou *et al.* 2016). Therefore, pretreatment using sodium hydroxide can be regarded as a feasible and effective method for lignocellulosic biomass processing.

Notwithstanding the advantages, determining the optimal combination of pretreatment parameters such as bark presence, steam duration, alkali concentration, and treatment time remains difficult due to the complexity and nonlinearity of their interactions. In this context, artificial intelligence, particularly artificial neural networks (ANNs), has emerged as a potent approach to model and optimize such multifactorial processes (Almeida 2002). ANNs have the capacity to learn complex, nonlinear relationships between input variables and response outputs without requiring explicit mathematical formulations. This renders them especially suitable for predicting outcomes in biological systems with inherent variability. Recent applications of artificial neural networks (ANN) in the field of biomass conversion have demonstrated a high degree of accuracy in modelling processes such as enzymatic saccharification, bioethanol yields, and the influence of pretreatment conditions (Vinitha *et al.* 2024; Azad *et al.* 2025). For instance, Vinitha *et al.* (2024) applied optimized decision-making algorithms to enhance the efficiency of enzymatic saccharification, while Azad *et al.* (2025) integrated orthogonal experimental designs with machine learning algorithms to achieve cellulose recovery efficiencies exceeding 88%, accompanied by minimal prediction error.

In line with these developments, the present study focused on pine biomass subjected to steam explosion and NaOH pretreatment. The efficiency of the enzymatic hydrolysis of such biomass is affected by multiple covariates, including the presence

of bark, the duration of steam explosion, the concentration of NaOH, and treatment time. The variables under consideration are pivotal in determining the chemical composition and physical structure of the resulting substrate. Consequently, it is challenging to predict hydrolysis performance using conventional methods alone.

In order to address this issue, a novel approach is proposed, which involves the use of an artificial neural network (ANN)-based modelling strategy to simulate and predict the efficiency of the enzymatic hydrolysis of pretreated pine biomass. The objective of training the model on experimental data is twofold: first, to analyze which process variables affect the enzymatic hydrolysis rate, and second, to evaluate the predictive performance of the ANN compared to other models such as random forest (RF), support vector regression (SVR), and extreme gradient boosting (XGB). Although machine learning algorithms have been increasingly applied in the fields of wood chemistry and bioprocessing, to the best of our knowledge, their application to modeling enzymatic hydrolysis based on the pretreatment of steam exploded pine wood biomass remains limited.

The objective of this study was to investigate a predictive model for enzymatic hydrolysis yield using pre-treatment process parameters of lignocellulosic biomass (pine wood biomass) such as steam explosion and alkaline pretreatment. An artificial neural network (ANN) was modeled using Python to predict the enzymatic hydrolysis rate based on these pretreatment variables. To enhance the prediction performance of the ANN, hyperparameter tuning was conducted. The optimized ANN model was then compared with other machine learning models in terms of prediction accuracy. Predictive performance was evaluated using mean square error (MSE) and the coefficient of determination (R^2). Therefore, the results of this study are expected to provide valuable insight into the synergistic effect of steam explosion and alkaline pretreatment on the enzyme hydrolysis of pine wood, and contributes to the development of efficient conversion processes for producing high-value sugars from lignocellulosic biomass.

EXPERIMENTAL

Materials

The pine wood chips used in the study were sourced from Punglim Corporation (Daejeon, Korea). The pine wood chips were divided into bark and non-bark samples, and all samples were used in the experiments. The wood chips were chipped to a size of approximately 3 cm (W) × 3 cm (L) × 0.5 cm (H) and used for steam explosion treatment.

Pretreatment Process for Enzymatic Hydrolysis

Pine woodchip steam explosion

The steam explosion conditions used in this study followed the parameters described by Ha *et al.* (2024). The steam-explosion pretreatment process was conducted at the customized batch pilot unit (Yurim High Tech, Taegu, Gyeongsangbuk-do, Korea) based on the Masonite technology. The reactor was steam exploded with saturated steam at 25 kg/cm² for 1 to 5 min. The steam exploded pine wood chips were collected in circular bins, cooled, sealed in PE bags and refrigerated at 4 °C until chemical pretreatment. According to the report by Rodríguez *et al.* (2017), it has been demonstrated that relatively mild steam explosion conditions (pressure of 15 to 25 bar, temperature of 200 to 220 °C, and residence time of 1 to 5 minutes) can still enable high enzymatic hydrolysis yields.

Chemical pretreatment of steam explosion samples

The alkaline pretreatment was conducted using a modified version of the method described by Gunam *et al.* (2020). The steam exploded samples were pretreated with different concentrations of NaOH to increase the enzymatic hydrolysis rate. Different concentrations of NaOH (0.5-2%) were used for pretreatment. Sample and solvent were mixed in a 1:20 ratio in a 300 mL triangular flask. Pretreatment was carried out for 12 to 24 h at room temperature at rest. After pretreatment, the residue was filtered through Whatman filter paper No. 2 and washed with distilled water until the pH was neutral. The neutralized sample was used for enzymatic hydrolysis.

Evaluate the microstructure of chemical pretreatment samples

Scanning electron microscopy (SEM) (ZEISS Gemini 300, Germany) was used to evaluate the surface microstructure of steam explosion treated pine chips after chemical pretreatment. Samples were mounted on a stub using a conductive thermoplastic adhesive, coated with Pt on a Polaron E 5000 sputter coating unit and evaluated using scanning electron microscopy. The analysis was performed at an accelerating voltage of 5 Kv.

Enzymatic hydrolysis

The enzymatic hydrolysis was performed using a modified version of the method described by Bhalla *et al.* (2018). Chemically pretreated steam explosion samples were placed in 1 g in a 30 mL test tube for enzymatic hydrolysis and autoclaved at 121 °C for 30 min. After autoclaving, the sample was allowed to air dry on a clean bench. Buffer was prepared with Na-citrate buffer, 2% sodium azide, and Tween 80 (polysorbate 80). Ten mL of the buffer was added to a test tube and hydrolyzed by adding Cellic CTec3 (Novozymes, Denmark) 440FPU/glucan. After hydrolyzing at 50 °C, 210 rpm, for 72 hours, the hydrolysate was filtered through a 2G3 glass filter. After hydrolysis, the residue was dried at 105 ± 3 °C until constant weight and then weighed to calculate the rate of enzymatic hydrolysis.

Machine Learning and Optimization Approach

ANN modeling

The ANN was implemented using the TensorFlow package (Abadi *et al.* 2016), Keras package and Scikit-learn package of programs for learning from experimental data (Python 3. 10.6). Enzymatic degradation rates under different treatment conditions (steam explosion time, chemical pretreatment time, chemical concentration) were predicted using a multilayer back-propagation neural network. The network architecture has three input variables: steam explosion time (1, 3, 5 min), chemical pretreatment time (12h, 24h) and chemical concentration (0.5, 1, 2%). There was one output variable: enzymatic hydrolysis rate. All conditions were measured in triplicate, yielding a total of 108 data points. The data collected were subsequently divided into two equal segments: 80% was allocated for training, while the remaining 20% was designated for testing purposes. The model was trained using the K-fold validation method.

ANN hyperparameter optimization

A Keras tuner was used to optimize the hyperparameters of an artificial neural network (Saleh *et al.* 2022). This was done using Keras tuner randomsearchcv, randomizing the number of hidden layers, number of neurons, and dropout rate to find the values with the best model performance. The search ranges for each hyperparameter are summarized in Table 1.

Table 1. Scope of Hyperparameter Exploration in ANN Models

Hyperparameters	Range
Number of Hidden Layers	1 to 3
Units per Layer	8 to 64
Dropout Rate	0.0 to 0.3
Optimizer	Adam, Rmsprop, SGD
Learning rate	0.01, 0.001, 0.0001
Batchnorm	True, False

Adjustments were made of the number of hidden layers (1 to 3), number of neurons per layer (8 to 64), and dropout rate (0 to 0.3). The optimizer of choice was Adam, Rmsprop, and SGD, with learning rates of 0.01, 0.001, and 0.0001, and batch normalization set to True and False. The objective function was Mean Squared Error (MSE), and the hyperparameter values that minimize the MSE value were optimized by random search. To assess the efficacy of the model, a k-fold cross-validation was employed, with k set at 5.

Each combination was subjected to a training regime comprising 500 epochs. During this training, the loss value (MSE) of the validation dataset was meticulously monitored. Early termination of the training process was implemented if the value remained static for a period of 10 epochs.

Evaluating ANN models

To assess the comparative efficacy of ANN models, three machine learning algorithms were selected for analysis. Random forest (RF), support vector regression (SVR), and extreme gradient boosting (XGB) are algorithms frequently employed in regression and classification problems.

In order to determine the possibility of underfitting and overfitting of the ANN model, the data were split differently into train 75%, test 25% and train 85%, test 15%, and evaluated the ANN model for each data split.

The error can be calculated by comparing the difference between the predicted value and the target value. To minimize the error in the backpropagation algorithm, the weights and bias values of the previous layer and the backpropagated error were readjusted. This iterative process enhances the sequential model's performance, ensuring precise predictions for novel instances. The error evaluation criteria considered are mean square error (MSE) and coefficient of determination (R^2), defined as follows (Chicco *et al.* 2021),

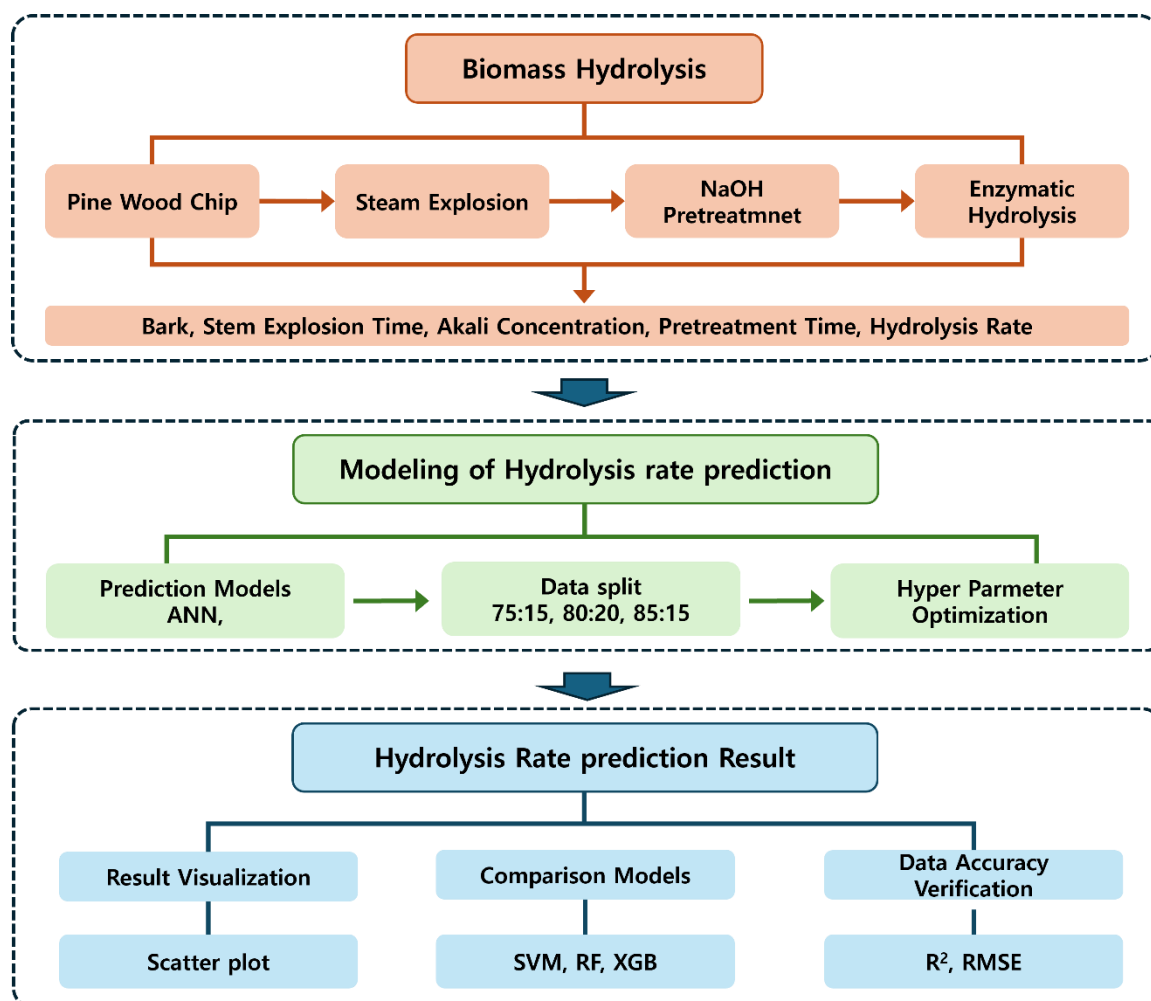


Fig. 1. Analytical workflow for modeling the prediction of enzymatic hydrolysis rate

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{predicted} - y_{actual})^2 \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{predicted} - y_{actual})^2}{\sum_{i=1}^n (y_{predicted} - y_{mean})^2} \quad (2)$$

where n , $y_{predicted}$, y_{actual} , and y_{mean} are the number of instances, the values generated by the ANN model, the target value, and the average value of the target output, respectively.

The modeling for predicting enzymatic hydrolysis yield was carried out through data acquisition, model optimization, and comparison with other machine learning algorithms. The overall workflow and methodology adopted in this study are illustrated in Fig. 1. This workflow may serve as a guideline for reproducing the results or applying similar approaches in related studies.

RESULTS AND DISCUSSION

Steam Explosion and Alkaline Pretreatment of Pine Wood Chips

As illustrated in Fig. 2, the original sample and the pine wood chip containing bark are represented visually. As the intensity of the steam blasting treatment

increased, a shift in color towards darker shades was observed, which was consistent across both the pine biomass with bark and the biomass without bark.

NaOH was chosen for the chemical pretreatment. The morphological alterations in pine biomass that had undergone pretreatment with NaOH were observed by means of SEM, with the resultant images presented in Fig. 3. The SEM images obtained demonstrated an augmentation in surface area and a disruption of cell walls as a consequence of chemical pretreatment.

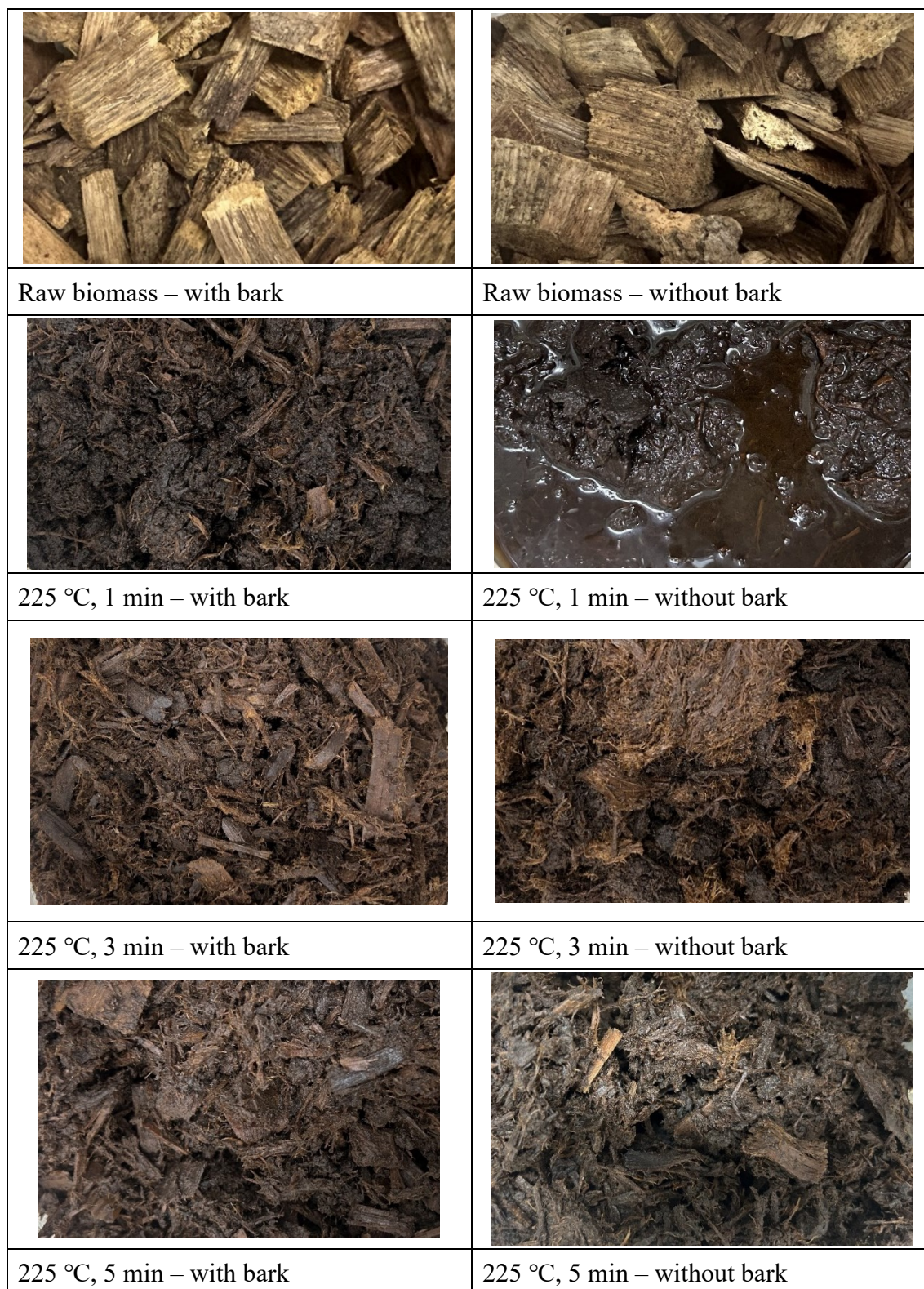


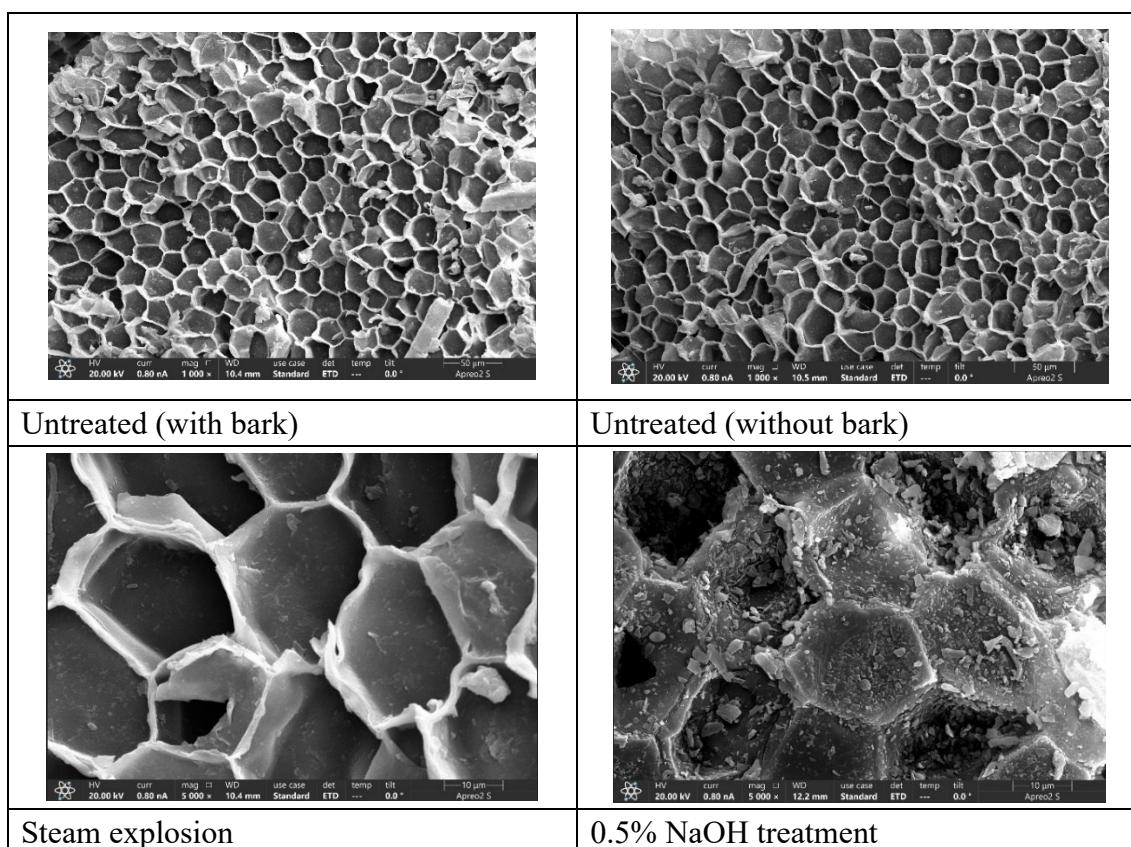
Fig. 2. Images of pine wood chip before and after steam explosion pretreatment

It was visually evident that the higher the solvent concentration, the more cell wall disruption occurred. However, it was found that sufficient cell wall disruption occurred even at a 2% NaOH concentration and room temperature, which is relatively lower than the typical NaOH treatment concentration. This indicates that treating biomass with low concentrations of chemical and mild temperatures can be effective in increasing biomass digestibility (Lou *et al.* 2016).

Enzymatic Hydrolysis of Biomass and the Influence of Process Variables

Before using an artificial neural network to train data, the scientific basis for the input variables must be clarified. The presence of bark is important because it can worsen the efficiency of enzymatic hydrolysis (Kim *et al.* 2005). During the steam blasting process, woody cellulosic biomass undergoes self-hydrolysis due to heat and steam, and fiber rupture due to overpressure, which facilitates enzyme penetration (Jacquet *et al.* 2010). Moreover, the degree of self-hydrolysis and fiber rupture varies depending on the time that the steam reacts with the wood cellulosic biomass, which is an important variable (Jacquet *et al.* 2012). Furthermore, the alkali concentration and chemical pretreatment time used in the chemical pretreatment may result in different enzymatic hydrolysis efficiencies due to different removal of enzymatic hydrolysis inhibitors (Persson *et al.* 2002).

In Table S1 and Fig. 4, the code “1” denotes including bark and “0” denotes not including bark. Samples without bark(0) had a higher median value than samples with bark(1), while samples with bark contained several outliers and showed a large variance. Chemical pretreatment time showed high hydrolysis rates for the 24h treatment and relatively few outliers. Chemical concentrations showed the highest enzymatic hydrolysis rates at 1%, and overall low hydrolysis rates (54.9 to 89.7%) at 0.5%.



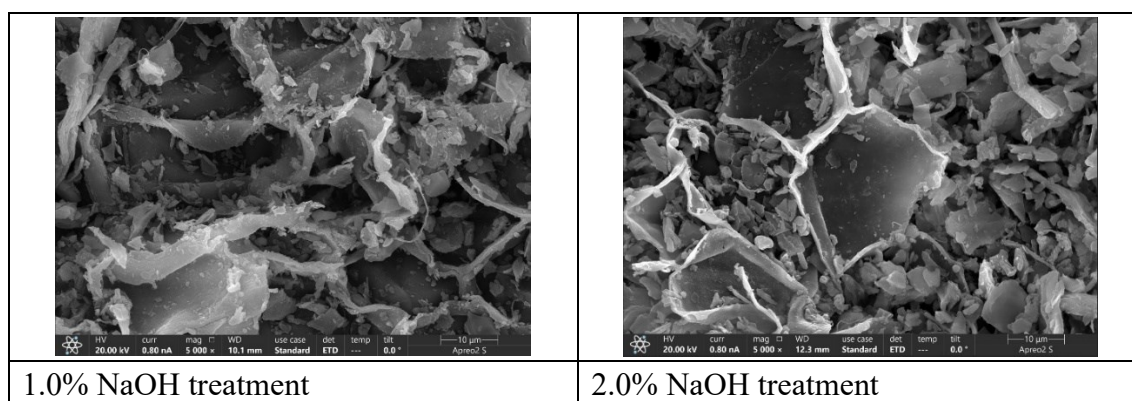


Fig. 3. Scanning electron microscopy (SEM) images of pine wood biomass obtained from untreated, steam explosion treatment and NaOH treatment

Steam explosion time had the highest median and least dispersion in the 5 min treated samples. The presence or absence of bark in the steam exploded samples did not appear to have a significant effect on enzymatic hydrolysis rates, which is consistent with previous reports (Kim et al. 2005). It can be seen from Fig. 1 that the longer the steam explosion time, the more structure destruction occurred due to the longer reaction time between the steam and biomass. It has been shown that broken-down biomass allows easier penetration of the enzyme and increases the hydrolysis rate (Jacquet et al. 2012). Therefore, a complex learning network should be built to predict the enzymatic hydrolysis rate by examining the correlation between the input variables. Figure 4 shows the correlation between the input and output variables.

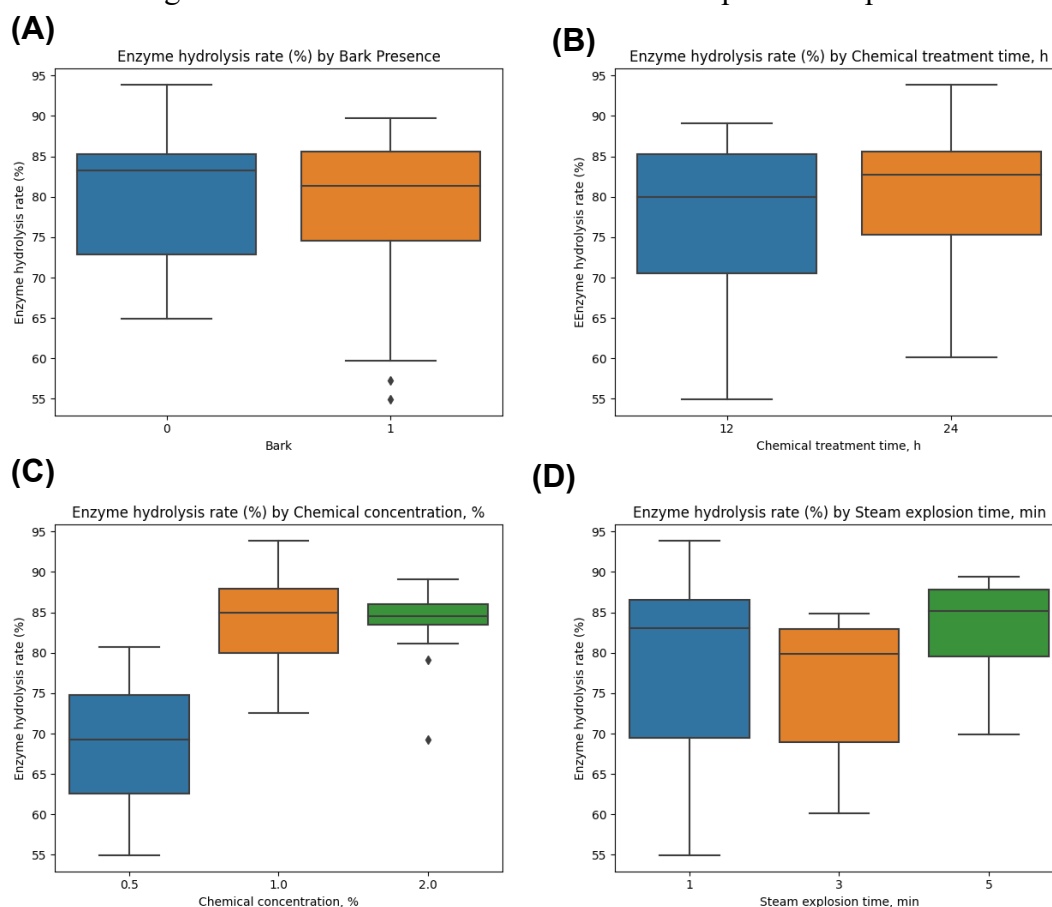


Fig. 4. The following box plots illustrate the effect of varying parameters on the efficiency of enzymatic hydrolysis. The parameters under investigation are as follows: bark (A), chemical pretreatment time (B), chemical concentration (C), and steam explosion time (D)

Figure 5 presents a correlation matrix analysis that includes the relationships among bark presence, steam explosion time, chemical concentration, chemical treatment time, and enzymatic hydrolysis yield. The matrix values indicate both the direction and strength of the relationships between variables. According to the correlation analysis, chemical concentration exhibited a strong positive correlation (0.61) with enzymatic hydrolysis yield, suggesting that increases or decreases in chemical concentration directly affect saccharification performance.

In contrast, steam explosion time and chemical treatment time showed very weak correlations with enzymatic hydrolysis yield. Additionally, the presence of bark demonstrated a weak negative correlation (-0.14) with enzymatic hydrolysis yield. Mild alkaline pretreatment followed by enzymatic hydrolysis has been reported to significantly enhance sugar yield in popular. Similarly, in non-woody lignocellulosic materials such as corn stover and rice straw, mild NaOH pretreatment has been shown to facilitate enzymatic hydrolysis (Ioelovich and Morag 2012).

ANN Modeling for Enzymatic Hydrolysis Rate Prediction

In this study, an artificial neural network (ANN) model was employed to predict the enzymatic hydrolysis rates from data exhibiting a nonlinear relationship. To prevent model overfitting and enhance its generalization performance, the dataset was randomly partitioned into training and testing subsets, with 80% allocated for training and the remaining 20% for testing.

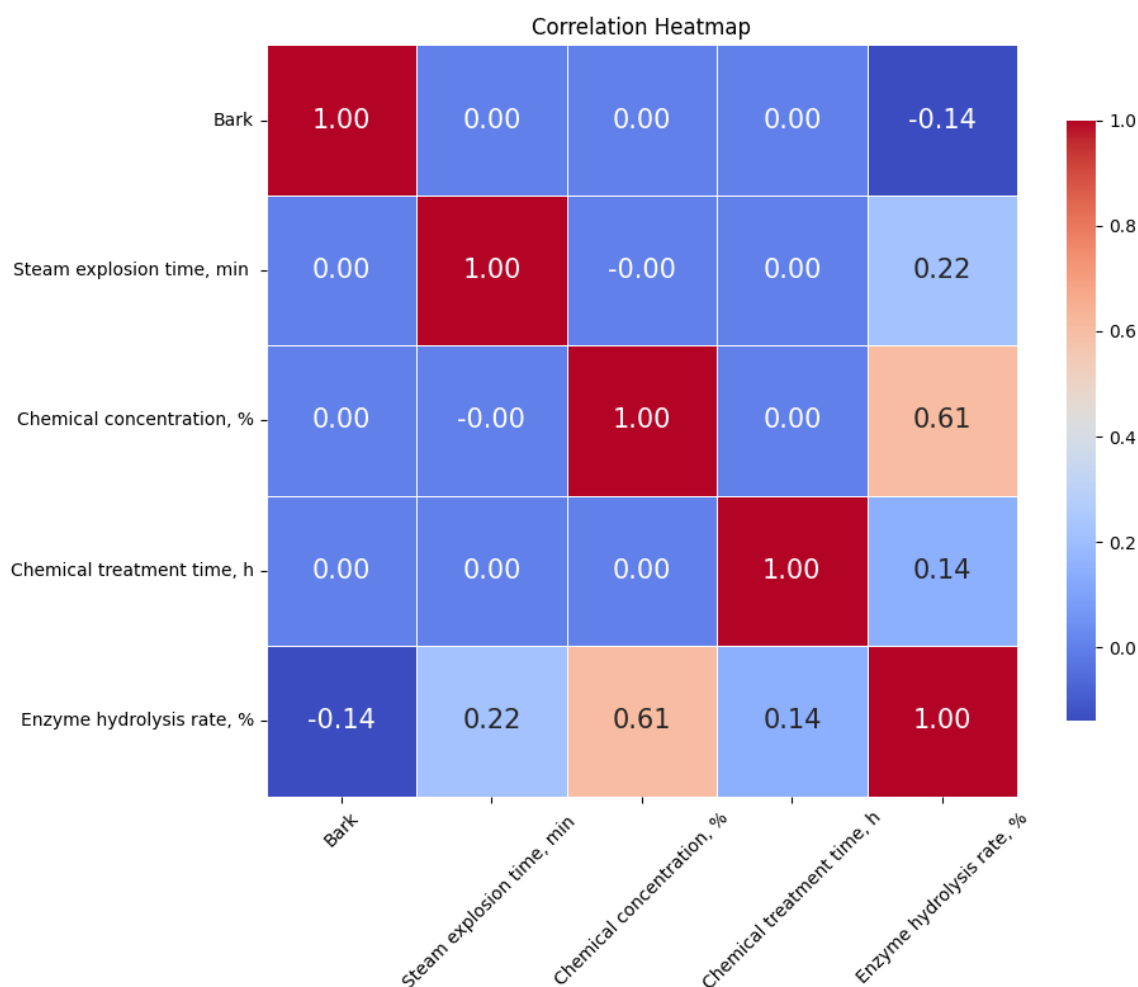


Fig. 5. Pearson correlation coefficient to analyses the relationship between process variables and enzymatic hydrolysis rates

Hyperparameter optimization of the ANN model was conducted using the random search method. This approach has been reported to yield optimal models more efficiently compared to grid search methods, either achieving similar performance or requiring less computational time (Bergstra and Bengio 2012). Multiple ANN training iterations were performed to identify hyperparameters that minimized the loss function, mean squared error (MSE). As illustrated in Fig. 6, the MSE value converged to a minimum and remained constant after approximately 175 training runs, indicating stabilization of the ANN model parameters.

The optimized ANN model architecture consisted of an input layer that integrated four process variables: presence of bark, chemical pretreatment time, chemical concentration, and steam explosion time. This input layer was connected to a hidden layer with 64 neurons and a single output layer that produced the predicted enzymatic hydrolysis yield. Each neuron computed the weighted sum of the inputs, added a bias term, and passed the result to the next layer.

The Rectified Linear Unit (ReLU) function was employed as the activation function for the hidden layer. Dropout was set to 0, and batch normalization was disabled (Batchnorm = False). The Adam optimizer was used to iteratively adjust the network weights based on the computed loss function. The architecture of the optimized model is illustrated in Fig. 7.

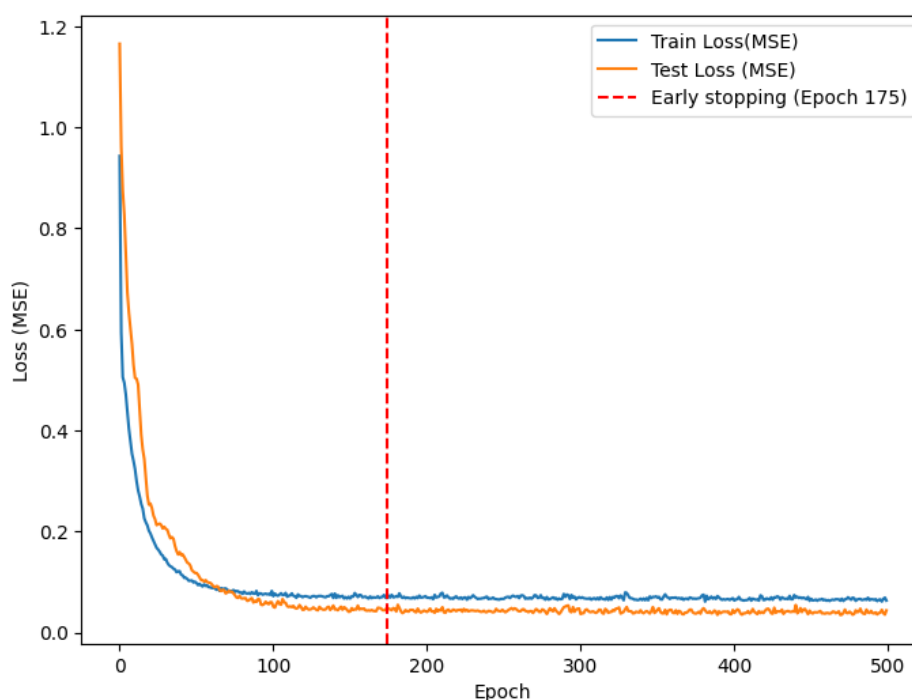


Fig. 6. Mean square error (MSE) of the proposed ANN model for predicting the enzyme hydrolysis rate of 500 epochs

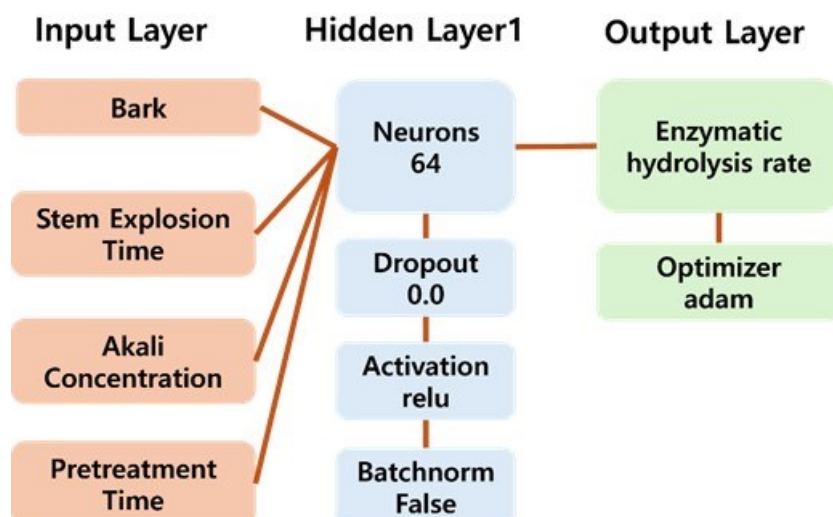


Fig. 7. Neural network topology of the optimized ANN model

To evaluate the potential for underfitting and overfitting in the optimized model, the model was trained and assessed using different data split ratios. (Table 2)

Table 2. Performance of Different data Split Ratio for Prediction of Enzyme Hydrolysis Rate

Data Split Ratio (Train: Test)	Enzyme Hydrolysis Rate			
	Train R ²	Train MSE	Test R ²	Test MSE
75:25	0.9331	4.4124	0.9781	2.5552
80:20	0.9373	4.4100	0.9805	2.4258
85:15	0.9381	4.4053	0.9815	2.3417

As shown in Table 2, when the data split ratio was adjusted to 85:15 using the same dataset, the resulting Train R² and Test R² were 0.9381 and 0.9815, respectively, and the Train MSE and Test MSE were 4.4053 and 2.3417. Compared to the results obtained with an 80:20 split, the variation in error metrics remained within 3%, indicating that overfitting was not observed even when the proportion of training data increased. Likewise, under a 75:25 split ratio, the Train R² and Test R² values were 0.9331 and 0.9781, respectively, and the Train and Test MSE values were 4.7824 and 2.5552. The performance deviation remained within 5% relative to the 80:20 baseline, suggesting that underfitting did not occur when the training data size was reduced.

These results validate that the developed artificial neural network (ANN) model effectively learned the data without experiencing underfitting or overfitting, and it can robustly predict enzymatic hydrolysis yield based on process parameters. The train-test split ratio is known to be a critical factor that influences model accuracy (Huang *et al.* 2023). Many researchers typically follow the Pareto principle and adopt an 80:20 data split for model development (Chen *et al.* 1994). The present study also supports the suitability of this principle, as evidenced by the results in Table 2.

Model Comparison Analysis

Both graphical and statistical approaches were employed to compare the predictive capabilities of the ANN, RF, SVR, and XGB models with respect to the enzymatic hydrolysis rate of steam-exploded pine wood chips. The accuracy of each model was evaluated using two statistical metrics: the coefficient of determination

(R^2) and the mean squared error (MSE). As presented in Table 3, the training dataset yielded high R^2 values of 0.9373, 0.9226, 0.9250, and 0.8260 for ANN, RF, SVR, and XGB models, respectively. Correspondingly, low MSE values of 4.4100, 5.4486, 5.2497, and 12.2497 were obtained for ANN, RF, SVR, and XGB models, respectively. On the testing dataset, the ANN model exhibited the highest R^2 value (0.9805) and the lowest MSE (2.4258), followed by SVR ($R^2 = 0.9463$, MSE = 3.1451), RF ($R^2 = 0.9208$, MSE = 8.4556), and XGB ($R^2 = 0.8674$, MSE = 14.1451).

Table 3. Performance of Different Modeling Methods for Prediction of Enzyme Hydrolysis Rate

Modeling Approaches and Statistical Parameters	Enzyme Hydrolysis Rate			
	Train R^2	Train MSE	Test R^2	Test MSE
ANN	0.9373	4.4100	0.9805	2.4258
RF	0.9226	5.4486	0.9208	8.4556
SVR	0.9250	5.2497	0.9463	3.1451
XGB	0.8260	12.2497	0.8674	14.1451

Among the evaluated models, ANN, RF, and SVR exhibited excellent predictive performance, with ANN clearly demonstrating superior accuracy in predicting the enzymatic hydrolysis rate. The lower MSE values for the ANN model, especially on external validation samples, confirmed its robustness and strong predictive capacity. These results validate that the enzymatic hydrolysis rate of pine biomass can be reliably and accurately estimated using ANN modeling based on the selected process variables. The coefficient of determination (R^2) measures the accuracy of predictions relative to observed target values and is frequently utilized to assess model performance.

Figure 8 compares the R^2 values obtained from each model. All models demonstrated strong predictive performance, yielding R^2 values above 0.93 for both training and testing datasets. Notably, the ANN model exhibited superior predictive capability for estimating enzymatic hydrolysis rates compared to the RF, SVR, and XGB models, achieving the highest R^2 value of 0.97 on the test dataset. These findings align with previous studies that have highlighted the strong predictive capacity of ANN models in biomass conversion processes (Persson *et al.* 2002; Vibha *et al.* 2024; Jayakumar *et al.* 2025).

Vinitha *et al.* (2023) applied a machine learning approach using decision tree algorithms to optimize the enzymatic hydrolysis of biomass. Their study employed process variables as training data and reported a high coefficient of determination ($R^2 = 0.9762$), indicating strong predictive performance. Similarly, Khangwal *et al.* (2021) introduced a multilayer feed-forward artificial neural network model to predict sugar production from hemicellulose extracted from corn cobs. Using process variables as model inputs, they achieved an R^2 of 0.9651, which is comparable to the performance of our ANN model. De Farias Silva *et al.* (2022) investigated the use of both artificial neural networks and support vector machines to predict fermentation yield from *Sargassum* macroalgae. Their results showed that the ANN model with 15 neurons achieved an R^2 of 0.877, while the SVM model using a polynomial kernel function reached an R^2 of 0.821. These findings suggest that ANN models offer superior performance over SVM in yield prediction tasks, which is consistent with our study. To the best of our knowledge, there has been no prior research applying machine learning algorithms to predict the enzymatic hydrolysis yield of steam-

exploded pine pretreated under mild NaOH conditions. This highlights the novelty and contribution of our work to the field of lignocellulosic biomass valorization.

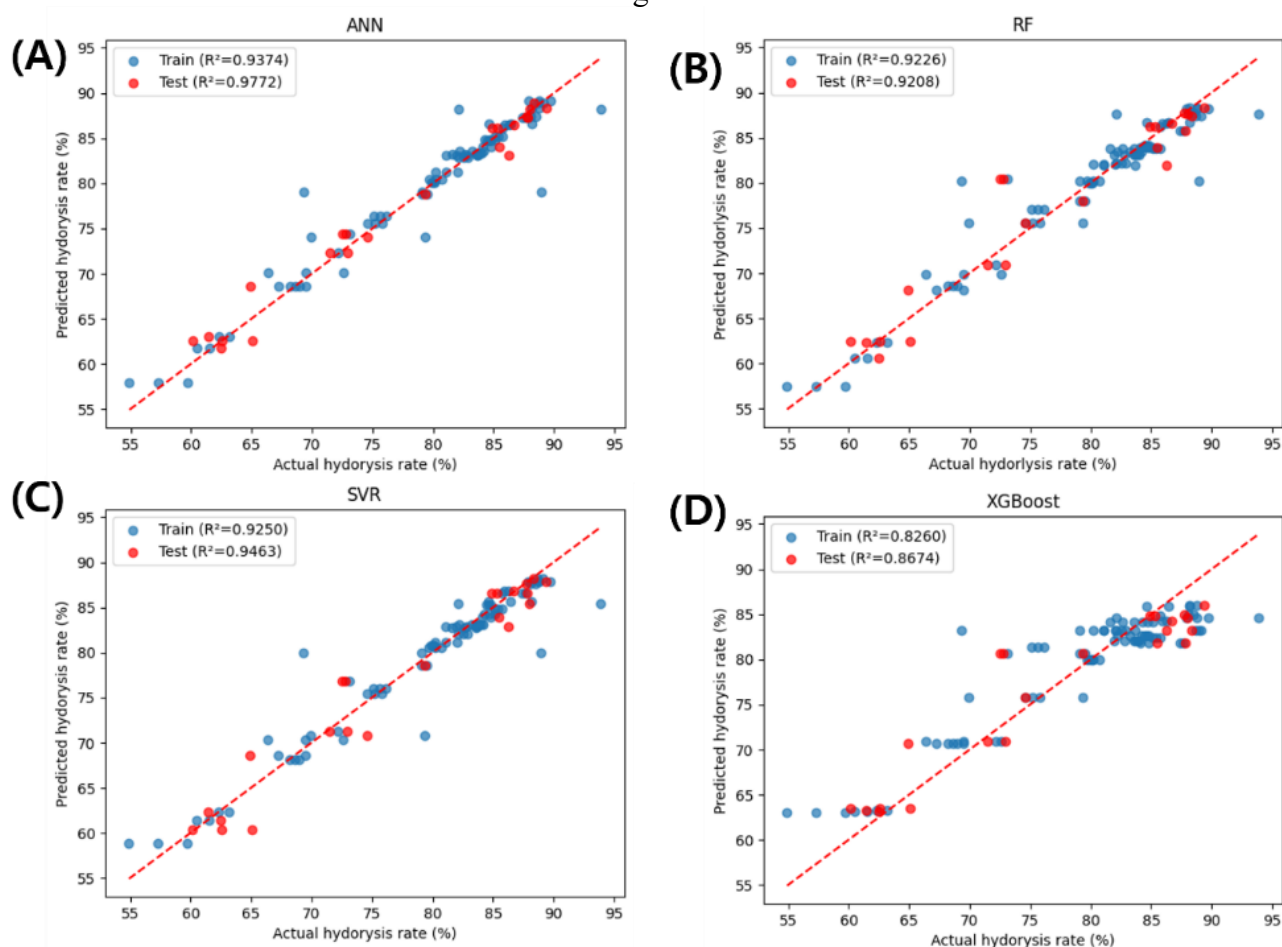


Fig. 8. Comparison between the actual and predicted enzymatic hydrolysis rates for artificial neural networks (ANN), random forests (RF), support vector regression (SVR), and XGB models (D)

CONCLUSIONS

1. This study demonstrated that an artificial neural network (ANN) model can successfully predict the enzymatic hydrolysis rate of steam-exploded pine wood chips subjected to mild alkaline pretreatment. The ANN model outperformed conventional models such as random forest (RF), support vector regression (SVR), and XGB, achieving an R^2 of 0.9805 and a mean squared error (MSE) of 2.43 on the test dataset.
2. Among the evaluated process parameters, the chemical concentration, particularly at 1% NaOH, exhibited the strongest linear correlation ($r = 0.61$) with the enzymatic hydrolysis rate. Conversely, chemical pretreatment time and steam explosion duration demonstrated significant yet nonlinear impacts on hydrolysis performance. Optimal enzymatic hydrolysis conditions were determined to be 1% NaOH concentration, a 24-hour pretreatment duration, 5 minutes of steam explosion treatment, and the absence of bark, achieving a maximum enzymatic hydrolysis rate of 93.9%.

3. This study confirms that artificial neural networks (ANNs) effectively model the complex nonlinear relationships inherent in biomass hydrolysis processes. Thus, ANN provides a robust, rapid, and precise predictive approach, facilitating the optimization of biomass bioconversion conditions.

ACKNOWLEDGMENTS

This study was carried out with the support of R&D Program for Forest Science Technology (Project No. "RS-2023-KF00245261382116530003") provided by Korea Forest Service (Korea Forestry Promotion Institute).

Data Availability

All datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

REFERENCES CITED

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., and Isard, M. (2016). "TensorFlow: A system for large-scale machine learning," *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pp. 265-283.
- Almeida, J. S. (2002). "Predictive non-linear modeling of complex data by artificial neural networks," *Curr. Opin. Biotechnol.* 13(1), 72-76. DOI: 10.1016/S0958-1669(02)00288-4
- Antonopoulou, G., Vayenas, D., and Lyberatos, G. (2016). "Ethanol and hydrogen production from sunflower straw: The effect of pretreatment on the whole slurry fermentation," *Biochem. Eng. J.* 116, 65-74. DOI: 10.1016/j.bej.2016.06.014
- Azad, S. A., Madadi, M., Rahman, A., Sun, C., and Sun, F. (2025). "Machine learning-driven optimization of pretreatment and enzymatic hydrolysis of sugarcane bagasse: Analytical insights for industrial scale-up" *Fuel* 390, article 134682. DOI: 10.1016/j.fuel.2025.134682
- Bergstra, J., and Bengio, Y. (2012). "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.* 13(1), 281-305.
- Bhalla, A., Bansal, N., Pattathil, S., Li, M., Shen, W., Particka, C. A., Karlen, S. D., Phongpreecha, T., Semaan, R. R., Gonzales-Vigil, E., Ralph, J., Mansfield, S. D., Ding, S., Hodge, D. B., and Hegg, E. L. (2018). "Engineered lignin in poplar biomass facilitates Cu-catalyzed alkaline-oxidative pretreatment," *ACS Sustainable Chem. Eng.* 6(3), 2932–2941. DOI: 10.1021/acssuschemeng.7b02067
- Cai, C., Zhang, C., Li, N., Liu, H., Xie, J., Lou, H., Pan, X., Zhu, J. Y., and Wang, F. (2023). "Changing the role of lignin in enzymatic hydrolysis for a sustainable and efficient sugar platform," *Renew. Sustain. Energy Rev.* 183, article 113445. DOI: 10.1016/j.rser.2023.113445
- Chen, Y., Chong, P. P., and Tong, M. Y. (1994). "Mathematical and computer modelling of the Pareto principle," *Math. Comput. Model.* 19(9), 61–80. DOI: 10.1016/0895-7177(94)90041-8

- Chicco, D., Warrens, M. J., and Jurman, G. (2021). "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.* 7, article e623. DOI: 10.7717/peerj-cs.623
- De Farias Silva, C. E., Costa, G. Y. S. C. M., Ferro, J. V., Carvalho, F. O., da Gama, B. M. V., Meili, L., Silva, M. C. S., Almeida, R. M. R. G., and Tonholo, J. (2022). "Application of machine learning to predict the yield of alginate lyase solid-state fermentation by *Cunninghamella echinulata*: Artificial neural networks and support vector machine," *React. Kinet. Mech. Catal.* 135(6), 3155-3171. DOI: 10.1007/s11144-022-02293-9
- Ding, S., and Himmel, M. E. (2006). "The maize primary cell wall microfibril: A new model derived from direct visualization," *J. Agric. Food Chem.* 54(3), 597-606. DOI: 10.1021/jf051851z
- Gunam, I., Setiyo, Y., Antara, N. S., Wijaya, I., Arnata, I. W., and Putra, I. (2020). "Enhanced delignification of corn straw with alkaline pretreatment at mild temperature," *Rasayan J. Chem.* 13, 1022-1029. DOI: 10.31788/RJC.2020.1325573
- Ha, S. Y., Jung, J. Y., Kim, H. C., Lim, W. S., and Yang, J. (2024). "Low-temperature and low-concentration sodium hydroxide pretreatment for enhanced enzyme hydrolysis rate from *Quercus variabilis* Blume," *BioResources* 19(2), 2592-2604. DOI: 10.15376/biores.19.2.2592-2604
- Huang, F., Teng, Z., Guo, Z., Catani, F., and Huang, J. (2023). "Uncertainties of landslide susceptibility prediction: Influences of different spatial resolutions, machine learning models and proportions of training and testing dataset," *Rock Mech. Bull.* 2(1), article 100028. DOI: 10.1016/j.rockmb.2023.100028
- Ioelovich, M., and Morag, E. (2012). "Study of enzymatic hydrolysis of mild pretreated lignocellulosic biomasses," *BioResources* 7(1), 1040-1052. DOI: 10.15376/biores.7.1.1040-1052
- Jacquet, N., Vanderghem, C., Danthine, S., Quiévy, N., Blecker, C., Devaux, J., and Paquot, M. (2012). "Influence of steam explosion on physicochemical properties and hydrolysis rate of pure cellulose fibers," *Bioresource Technol.* 121, 221-227. DOI: 10.1016/j.biortech.2012.06.073
- Jacquet, N., Vanderghem, C., Blecker, C., and Paquot, M. (2010). "La steam explosion: Application en tant que prétraitement de la matière lignocellulosique," *BASE* 14(2), 118-128.
- Jayakumar, M., Thiyagar, T., Abo, L. D., Arumugasamy, S. K., and Jabesa, A. (2025). "Paddy straw as a biomass feedstock for the manufacturing of bioethanol using acid hydrolysis and parametric optimization through response surface methodology and an artificial neural network," *Biomass Convers. Biorefin.* 15(3), 3803-3825. DOI: 10.1007/s13399-024-05371-1
- Kawaguchi, H., Hasunuma, T., Ogino, C., and Kondo, A. (2016). "Bioprocessing of bio-based chemicals produced from lignocellulosic feedstocks," *Curr. Opin. Biotechnol.* 42, 30-39. DOI: 10.1016/j.copbio.2016.02.031
- Khangwal, I., Chhabra, D., and Shukla, P. (2021). "Multi-objective optimization through machine learning modeling for production of xylooligosaccharides from alkali-pretreated corn-cob xylan via enzymatic hydrolysis," *Indian J. Microbiol.* 61(4), 458-466. DOI: 10.1007/s12088-021-00970-2
- Kim, K. H., Tucker, M., and Nguyen, Q. (2005). "Conversion of bark-rich biomass mixture into fermentable sugar by two-stage dilute acid-catalyzed hydrolysis," *Bioresource Technol.* 96(11), 1249-1255. DOI: 10.1016/j.biortech.2004.10.017

- Kim, J. S., Lee, Y. Y., and Kim, T. H. (2016). "A review on alkaline pretreatment technology for bioconversion of lignocellulosic biomass," *Bioresour. Technol.* 199, 42–48. DOI: 10.1016/j.biortech.2015.08.08
- Kumar, B., Bhardwaj, N., Agrawal, K., Chaturvedi, V., and Verma, P. (2020). "Current perspective on pretreatment technologies using lignocellulosic biomass: An emerging biorefinery concept," *Fuel Process. Technol.* 199, article 106244. DOI: 10.1016/j.fuproc.2019.106244
- Lou, H., Hu, Q., Qiu, X., Li, X., and Lin, X. (2016). "Pretreatment of miscanthus by NaOH/urea solution at room temperature for enhancing enzymatic hydrolysis," *Bioenergy Res.* 9(1), 335–343. DOI: 10.1007/s12155-015-9695-x
- Nges, I. A., Li, C., Wang, B., Xiao, L., Yi, Z., and Liu, J. (2016). "Physio-chemical pretreatments for improved methane potential of *Miscanthus lutarioriparius*," *Fuel* 166, 29–35. DOI: 10.1016/j.fuel.2015.10.108
- Persson, P., Andersson, J., Gorton, L., Larsson, S., Nilvebrant, N., and Jönsson, L. J. (2002). "Effect of different forms of alkali treatment on specific fermentation inhibitors and fermentability of lignocellulose hydrolysates," *J. Agric. Food Chem.* 50(19), 5318–5325. DOI: 10.1021/jf025565o
- Rodríguez, F., Sanchez, A., and Parra, C. (2017). "Role of steam explosion on enzymatic digestibility, xylan extraction, and lignin release of lignocellulosic biomass," *ACS Sustainable Chem. Eng.* 5(6), 5234–5240. DOI: 10.1021/acssuschemeng.7b00580
- Saleh, H., Hussien, A. M., Hassan, M. R., and Ali, A. A. (2022). "Predicting stroke disease based on recurrent neural network and optimization techniques," 2022 *International Conference on Engineering & MIS (ICEMIS)*, 1–5. DOI: 10.1109/ICEMIS56295.2022.9914334
- Vibha, R., Sandesh, K., Ujwal, P., and Shet, V. B. (2024). "RSM- and ANN-based modeling for a novel hydrolysis process of lignocellulose residues to produce cost-effective fermentable sugars," *Biomass Convers. Biorefin.* 14(19), 24181–24196. DOI: 10.1007/s13399-023-04484-3
- Vinitha, N., Vasudevan, J., and Gopinath, K. P. (2023). "Bioethanol production optimization through machine learning algorithm approach: Biomass characteristics, saccharification, and fermentation conditions for enzymatic hydrolysis," *Biomass Convers. Biorefin.* 13(8), 7287–7299. DOI: 10.1007/s13399-022-03163-z
- Vinitha, N., Vasudevan, J., Gopinath, K. P., Arun, J., Madhu, S., and Naveen, S. (2024). "Enhancing the dilute acid hydrolysis process using a machine learning approach: Investigation of different biomass feedstocks influences glucose and ethanol yields," *Biomass Convers. Biorefin.* 15, 9159–9171. DOI: 10.1007/s13399-024-05714-y
- Vu, H. P., Nguyen, L. N., Vu, M. T., Johir, M. A. H., McLaughlan, R., and Nghiem, L. D. (2020). "A comprehensive review on the framework to valorise lignocellulosic biomass as biorefinery feedstocks," *Sci. Total Environ.* 743, article 140630. DOI: 10.1016/j.scitotenv.2020.140630
- Yu, Z., Zhang, B., Yu, F., Xu, G., and Song, A. (2012). "A real explosion: The requirement of steam explosion pretreatment," *Bioresour. Technol.* 121, 335–341. DOI: 10.1016/j.biortech.2012.06.055
- Zabed, H. M., Akter, S., Yun, J., Zhang, G., Awad, F. N., Qi, X., and Sahu, J. N. (2019). "Recent advances in biological pretreatment of microalgae and lignocellulosic biomass for biofuel production," *Renew. Sustain. Energy Rev.* 105, 105–128. DOI: 10.1016/j.rser.2019.01.048

Zhang, Q., Wan, Z., Yu, I. K. M., and Tsang, D. C. W. (2021). "Sustainable production of high-value gluconic acid and glucaric acid through oxidation of biomass-derived glucose: A critical review," *J. Clean. Prod.* 312, 127745. DOI: 10.1016/j.jclepro.2021.127745

Article submitted: May 28, 2025; Peer review completed: July 1, 2023; Revised version received: July 21, 2025; Accepted: July 25, 2025; Published: August 1, 2025. DOI: 10.15376/biores.20.4.8400-8419

APPENDIX

Table S1. Enzyme Hydrolysis Rate of Pine Wood Biomass as Various NaOH Treatment Conditions

Bark	Steam explosion time (min)	Chemical concentration (%)	Chemical treatment time (h)	Enzyme hydrolysis rate (%)
1	1	0.5	12	57.3
1	1	0.5	24	61.5
1	1	1.0	12	87.6
1	1	1.0	24	88.8
1	1	2.0	12	83.7
1	1	2.0	24	82.6
0	1	0.5	12	67.2
0	1	0.5	24	72.2
0	1	1.0	12	84.8
0	1	1.0	24	88.0
0	1	2.0	12	79.1
0	1	2.0	24	84.7
1	1	0.5	12	59.7
1	1	0.5	12	54.9
1	1	0.5	24	62.5
1	1	0.5	24	60.5
1	1	1	12	87.8
1	1	1	12	87.4
1	1	1	24	89.7
1	1	1	24	87.9
1	1	2	12	86.3
1	1	2	12	81.1
1	1	2	24	83.6
1	1	2	24	81.6
0	1	0.5	12	69.5
0	1	0.5	12	64.9
0	1	0.5	24	72.9
0	1	0.5	24	71.5
0	1	1	12	85.5
0	1	1	12	84.1
0	1	1	24	93.9
0	1	1	24	82.1
0	1	2	12	88.9
0	1	2	12	69.3
0	1	2	24	85.1
0	1	2	24	84.3
1	3	0.5	12	62.3
1	3	0.5	24	62.6
1	3	1	12	79.3
1	3	1	24	81.1
1	3	2	12	82.8
1	3	2	24	83.2
0	3	0.5	12	68.6
0	3	0.5	24	69.5
0	3	1	12	72.8
0	3	1	24	82.5
0	3	2	12	83.8
0	3	2	24	84.6
1	3	0.5	12	63.2
1	3	0.5	12	61.4
1	3	0.5	24	65.1
1	3	0.5	24	60.1
1	3	1	12	79.5
1	3	1	12	79.1
1	3	1	24	82.0
1	3	1	24	80.2
1	3	2	12	83.7

1	3	2	12	81.9
1	3	2	24	84.2
1	3	2	24	82.2
0	3	0.5	12	69.0
0	3	0.5	12	68.2
0	3	0.5	24	72.6
0	3	0.5	24	66.4
0	3	1	12	73.1
0	3	1	12	72.5
0	3	1	24	82.9
0	3	1	24	82.1
0	3	2	12	84.0
0	3	2	12	83.6
0	3	2	24	84.8
0	3	2	24	84.4
1	5	0.5	12	74.6
1	5	0.5	24	75.2
1	5	1	12	75.6
1	5	1	24	88.1
1	5	2	12	88.7
1	5	2	24	85.3
0	5	0.5	12	80.0
0	5	0.5	24	80.2
0	5	1	12	85.4
0	5	1	24	88.8
0	5	2	12	86.3
0	5	2	24	86.4
1	5	0.5	12	79.3
1	5	0.5	12	69.9
1	5	0.5	24	75.8
1	5	0.5	24	74.6
1	5	1	12	76.1
1	5	1	12	75.1
1	5	1	24	88.5
1	5	1	24	87.7
1	5	2	12	89.1
1	5	2	12	88.3
1	5	2	24	85.7
1	5	2	24	84.9
0	5	0.5	12	80.1
0	5	0.5	12	79.9
0	5	0.5	24	80.7
0	5	0.5	24	79.7
0	5	1	12	85.7
0	5	1	12	85.1
0	5	1	24	89.4
0	5	1	24	88.2
0	5	2	12	86.7
0	5	2	12	85.9
0	5	2	24	88.2
0	5	2	24	84.6