Industrial Crops & Products 206 (2023) 117686

Contents lists available at ScienceDirect

Industrial Crops & Products

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

Harnessing Artificial Neural Networks and large language models for bioprocess optimization: Predicting sugar output from Kraft waste-based lignocellulosic pretreatments

Anthea Naomi David [a], Y. Sewsynker-Sukai [b], E.L. Meyer [b], E.B. Gueguim Kana [a,*]

[a] University of KwaZulu-Natal, School of Life Sciences, Pietermaritzburg, South Africa
[b] University of Fort Hare, Fort Hare Institute of Technology, Private Bag X1314, Alice 5700, South Africa

ARTICLE INFO

Keywords:
Artificial Neural Network
Bioprocess modelling and optimization
Sensitivity analysis
Large language models
Kraft waste-based pretreatment

ABSTRACT

This study implements Artificial Neural Network (ANN) models as predictive tools for glucose responses from Kraft waste-based pretreatments. The developed steam- and microwave-assisted ANN models achieved R^2 scores > 0.95 for the observed and predicted glucose responses. An in-depth sensitivity analysis revealed that the glucose responses for the steam and microwave models were highly susceptible to the stepwise variation in green liquor dregs concentration (>3.3-fold) and power intensity (>2.6-fold), respectively. Comparative assessment on the capability of the large language model, ChatGPT, to generate innovative and factually accurate insights based on the process data was carried out. The novel process insights deduced by ChatGPT concurred with the authors' findings of this study, underscoring the unique critical role of integrating advanced artificial intelligence and domain-specific knowledge to accelerate progression in lignocellulosic waste pretreatment. As such, these synergies align with global sustainable developmental objectives that leverage 4IR technologies, propelling this research field forward.

1. Introduction

The overall global demand of fossil fuel-derived energy and commodity chemicals has caused widespread issues rooted in natural resource scarcity, global warming and waste accumulation. This has exacerbated the mandate to progress from a largely reliant non-renewable economy towards a carbon-neutral bioeconomy to achieve energy independence and effective management of the environmental footprint (Patel and Shah, 2021). In this respect, the conversion of waste-derived-lignocellulosic biomass to high value bioproducts has become an emerging concept in the race towards a more sustainable future. Lignocellulosic biomass (LCB), especially agricultural waste residues are energy-dense, non-food-based feedstocks that regenerates itself annually and in abundance (Nunes et al., 2020). This inexhaustible and low-cost feedstock contains cellulose (38–55%), hemicellulose (23–32%) and lignin (15–25%) that account for approximately 90% of the dry matter present within the substrate and may be converted into value added products such as biochemicals and bioenergy (Kim et al., 2015; Muthuvelu et al., 2019; McKendry, 2002). More specifically, LCB such as corn cob waste (CCW) and banana pseudostems (BPS) have been recognized for its abundant annual output, high growth rate, rich carbohydrate and low lignin content, making these substrates attractive precursors for high sugar recovery (David et al., 2021; Laltha et al., 2022). The amorphous hemicellulose networks link the cellulose chains into microfibrils and form crosslinks with the non-carbohydrate lignin polymers to provide mechanical strength and intense impermeability to the LCB (Rebello et al., 2020). As a result, pretreatment is necessary prior to bioconversion processes in order to destabilize the cohesive lignin structure for enhanced amenability to enzymatic reactions followed by microbial fermentation.

Alkaline pretreatment strategies have been tagged as a gold standard due to its high depolymerization activity, but the extensive operational costs and energy requirements have jeopardized industrial applicability (David et al., 2021). For this reason, a pretreatment technology that: (1) minimizes waste, (2) uses low toxicity processes and, (3) employs energy saving approaches while preserving the key production output will lead to a functional, economically sound lignocellulosic bioprocess. In an effort to streamline the search for an effective pretreatment method that meet these criteria, the spotlight has turned towards industrial waste derived pretreatment chemicals such as green liquor dregs (GLD) and

* Corresponding author.
E-mail address: kanag@ukzn.ac.za (E.B.G. Kana).

https://doi.org/10.1016/j.indcrop.2023.117686
Received 10 August 2023; Received in revised form 19 October 2023; Accepted 20 October 2023
Available online 1 November 2023

paper wastewater (PWW) as viable alternatives. Both GLD and PWW are waste residues generated in abundance from the Kraft paper and pulp mill industry that plague the system with filter blockages, hydraulic conductivity issues, extensive landfilling and wastewater treatment costs (David et al., 2020). However, its alkaline nature shows promise of greater traction over commonly reported laboratory-grade alkaline chemicals such as NaOH, NH_4OH and $NaHCO_3$ amongst others (David et al., 2020; David et al., 2021; Laltha et al., 2022). For instance, GLD contains alkaline species (Na_2CO_3 and Na_2S) that modify the structure of the cell wall matrix, while PWW exhibits dual functionality as a pretreatment catalyst and freshwater substitute (David et al., 2021). Repurposing these waste residues will minimize waste disposal, mitigate invasive biohazardous treatments, relieve the current water crisis, reduce pretreatment chemical costs and valorize waste constituents into marketable products.

In most cases, chemical catalysts have been associated with physical pretreatment strategies to enhance productivity and enable high recovery of fermentable sugars due to lignocellulosic degradation. In particular, steam-assisted heating ionizes water at high temperatures and pressures to catalyse the hydrolysis of ester and ether linkages (Toscan et al., 2019). Conversely, microwave irradiation, an energy-efficient mechanism uses direct heating technology to cause fibre fragmentation through molecular collisions (Sorn et al., 2019). Despite the mechanistic advantages of each of these processes, the system is still inundated with issues surrounding economics, productivity and bioprocess scale-up.

Lignocellulosic processes are influenced by cost factors related to raw material acquisition, resource usage, energy consumption, downstream processing, yield generation, product purification and market demand amongst others. Furthermore, achieving high conversion rates of LCBs into the desired products, maximizing product yields and minimizing product losses are vital to improve productivity. More especially, scaling up the lignocellulosic pretreatment processes from laboratory-scale to commercial production has long been a significant challenge since it involves maintaining process performance, productivity and product quality during its transition from small-scale reactors to larger industrial-scale equipment. Scale-up also impacts the overall economics of the process as capital costs, energy requirements, and operational considerations are continuously being modified. Addressing the aforementioned challenges encountered with lignocellulosic pretreatment processes requires a comprehensive understanding of the underlying science, optimization of bioprocess conditions, improved resource performance and integration of advanced process control strategies. These elements collectively contribute to the development of efficient and logically coordinated pretreatment technologies. By optimizing process parameters, improving efficiency and implementing effective scale-up strategies, the economics and productivity of lignocellulosic processes can be enhanced, leading to more sustainable and commercially viable biorefinery systems. As a result, these challenges necessitate the development of effective modelling and optimization strategies to address the research gap.

Modelling plays a pivotal role in determining the key conditions that will favourably impact the optimization of fermentable sugar outputs at the early stages of bioprocess development (Sindhu et al., 2016). The fourth industrial revolution (4IR) together with global sustainability goals have steered research in the direction of artificial intelligence (AI) systems such as Artificial Neural Networks (ANN), enabling virtual experiments that curb alarming costs and intensive labour. ANNs imitate the cognitive responses of the human brain to mathematically model complex non-linear systems without prior knowledge of kinetics, metabolic fluxes or the bioprocessing medium. Data-driven tools, such as ANNs, utilize correlations between the process inputs and corresponding outputs to train machine learning algorithms for accurate yield prediction, leading to maximized product generation and reduced inhibitors (Sebayang et al., 2017). Furthermore, ANNs employ feature selection methods to identify and rank critical input features within a dataset, thus minimizing redundancy, reducing training time, mitigating the risk of overfitting and facilitating data interpretation (Chen et al., 2020). Sensitivity analysis, a notable feature selection method, identifies vital input parameters for prediction of the output variable and quantifies how the changes in these input values affects the resultant output. Consequently, these intelligent models expedite screening assessments of the fermentable sugar release from pretreated LCB, offering foresight into industrial scale applications. Few studies have reported on the prediction of fermentable sugar yields from pretreated LCB using intelligent modelling. For example, Moodley et al. (2019) developed microwave- and steam- inorganic salt-based ANN models for the prediction of reducing sugar yields from sugarcane leaves using experimental data from published studies. The reported study revealed coefficient of determination (R^2) values of 0.97 for both models (Moodley et al., 2019). Similarly, Vani et al. (2015) applied ANN modelling to predict the effect of various input variables on the glucose and xylose concentrations during the enzymatic hydrolysis of rice straw. The high R^2 value of > 0.97 and error lying close to 0 indicates reliability of the ANN model in predicting the sugar yields during hydrolysis (Vani et al., 2015). In another instance, Lee et al. (2020) optimized an ultrasonic-assisted organosolv pretreatment process on oil palm empty fruit bunches for improved reducing sugar concentration. The Leverburgh Marquee ANN model indicated that ultrasonic-assisted organosolv pretreatment was significantly affected by temperature, time and sonication power with good predictive accuracies ($R^2 =0.91$) (Lee et al., 2020). In addition, Chang et al. (2011) subjected napiergrass to a two-stage pretreatment process of steam explosion (SE) followed by alkaline delignification. Three different methods were used, namely, Back-Propagation Neural Network (BPNN), the Multiple Linear Regression (MLR), and the Partial Least-Square regression (PLS) with enzymatic digestibility as the output. The data indicates that the BPNN model obtained a higher prediction performance ($R^2 >0.98$) compared to the MLR and PLS models (Chang et al., 2011). Rego et al. (2018) presented a study that compares the optimization of sugarcane bagasse delignification subjected to alkaline H_2O_2 using ANN and Adaptive Network based Fuzzy Inference System (ANFIS) with glucose and xylose concentrations as outputs. The statistical quality of the models was significant due to the low error values and high R^2 values close to 1 (Rego et al., 2018). As it stands, to the best of our knowledge, the application of ANN on Kraft waste-based pretreatment modelling has not yet been documented. Such intelligence approaches will provide real-time estimations of process parameters, enhance workflow by eliminating various laboratory tasks and negate experimentation costs.

While ANN is a robust tool for prediction, pattern recognition and complex decision-making, there has been a growing interest in exploring the potential of advanced large language models (LLMs), such as ChatGPT to elucidate complex perceptions in the domain of bioprocess development. By harnessing the capabilities of deep learning algorithms, the LLM possesses the ability to process, analyse and synthesize intricate patterns from extensive and multifaceted scientific data information, generating human-like articulations (Ray, 2023). Unlike contextual knowledge in human interpretation, LLMs provide unbiased analysis based on statistical associations presented in the dataset (Ray, 2023). However, these models lack domain-specific expertise and comprehension, potentially hindering its ability to provide nuanced interpretations intrinsic to the specialized field. To address this challenge, an integrative approach that synergistically combines the vast computational and data processing capabilities of LLMs with human interaction can be employed to ensure accurate and contextually informed decisions. This collaborative method aids in identifying potential discrepancies and cognitive biases that may arise in both human and model interpretations, ultimately propelling research advancements in factual, scientific knowledge generation.

Therefore, this study aims to develop two ANN models, steam- and microwave-assisted models, built on existing experimental data from previously published Kraft waste-based pretreatment regimes on lignocellulosic wastes (David, . et al., 2020; David et al., 2021; Laltha et al., 2022). The Kraft waste-based pretreatments included: (a) steam-assisted green liquor dregs (GLD) only in pure water (SGLD-W) (David et al., 2020), (b) steam-assisted combined GLD and paper wastewater (PWW) (SGLD-PWW) (David et al., 2021), (c) microwave-assisted combined GLD and PWW (MGLD-PWW) (David et al., 2021) and (d) microwave-assisted PWW (MPWW) (Laltha et al., 2022). Additionally, a sensitivity analysis was carried out to evaluate the impact of each input variable on the glucose yield output. These models will thereafter be able to identify data patterns and efficiently predict the glucose responses from new physicochemical input values that can be deployed as a virtual experimentation platform within a global public repository. In conjunction with ANN model development and contextual assessment, this study incorporates a unique element by introducing ChatGPT, a large language model (LLM) to the sensitivity analysis dataset. This explored the model's capability to recognize process data patterns, interpret data and generate factually sound insights in scientific reporting compared to already established human cognitive responses.

2. Materials and methods

2.1. Experimental data collection for ANN model development

The experimental data used for model development were obtained from our previous studies on Kraft waste-based pretreatment of corn cob waste (CCW) and banana pseudostem (BPS) (David et al., 2020; David et al., 2021; Laltha et al., 2022). These studies assessed the effect of various physicochemical parameters that enhanced the disintegration cross-linked lignin fractions of lignocellulosic waste for enzymatic conversion of cellulose to glucose. The pretreatments selected for model development were based on a: (a) steam-assisted green liquor dregs (GLD) only in pure water (SGLD-W) ((David et al., 2020), (b) steam-assisted combined GLD and paper wastewater (PWW) (SGLD-PWW) (David et al., 2021), (c) microwave-assisted combined GLD and PWW (MGLD-PWW) (David et al., 2021) and (d) microwave-assisted PWW (MPWW) (Laltha et al., 2022). A total of 96 experimental runs were used to develop the ANN model with the predictor being glucose yield. The selected input parameters and ranges for the microwave model included GLD concentration (0–50%), solid loading (10–30%), power intensity (100–900 W), heating time (1–10 min) and substrate type (corn cobs or banana pseudostem). For the steam model, the input variables consisted of GLD concentration (1–50%), solid loading (1–20%), heating temperature (80–121 °C), heating time (5–60 min) and water source (pure water or PWW). The resultant output for both models were glucose yield (g/g).

2.2. ANN model development for sugar yield prediction

Two ANN models, namely, steam-assisted and microwave-assisted Kraft waste-based models were structured on the Google Collaboratory platform using the Python version 2.4.1 (64-bit), with TensorFlow and Keras libraries. A feed-forward multi-layered perceptron (MLP) was used to model the non-linear relationship between the considered inputs and glucose yield from the Kraft waste pretreatments. The topology of both the steam- and microwave-assisted Kraft waste-based models consisted of one input layer to accommodate datasets with five predictor variables and two hidden layers with a dynamic number of nodes (20) within each layer (Table 1, Fig. 1A and B). With the aim of introducing non-linearity into the steam- and microwave-assisted Kraft waste-based models, the Rectified Linear Unit (ReLU) activation function was applied within these hidden layers. Lastly, the output layer consisted of a single node responsible for generating regression predictions. In particular, this layer applied no activation function, thus enabling the model to produce continuous numerical predictions without imposing any specific output limitations. For the feed forward training algorithm, the input data was scaled through weight assignment, which associated a real number quantity to the connection of two neurons. The neurons in the input layer transmitted the signals of the scaled data to the neurons in the hidden layer (Sewsynker-Sukai and Gueguim Kana, 2017a).

Table 1
Steam- and microwave-based model structure.

Layer (type)	Output shape	Number of parameters
Dense_256	None, 20	120
Dense_257	None, 20	420
Dense_258	None, 20	420
Dense_259	None, 1	21

2.3. ANN training and validation using backpropagation

Prior to using the data set, data normalization was carried out in the range [− 1 to − 1] according to Eq. (1).

$$\text{Normalized}(e_i) = \frac{e_i \quad -E_{min}}{E_{max} \quad -E_{min}} \tag{1}$$

where e_i is the normalized data and E_{min} and E_{max} denote the minimum and the maximum values.

Following data normalization, the input dataset was divided such that 80% of samples were assigned to the training set and the remaining 20% formed the validation set. Generally, the input layer processes the collated data and allocates each feature to an input node. The neurons are then stimulated through a forward-propagation approach such that the impact of each neuron's activation is regulated by the adjusted weights. These activators are propagated until the predicted output is achieved (Fig. 2). Subsequently, the back propagation algorithm employed to train the model adjusted the neuron's connections until the net error between the experimental and predicted data was reduced to a level below the target threshold. The weights were then revised based on their contribution to the error. The training process was repeated until the minimum mean square error was attained. A grid search approach was employed to optimize the model's hyperparameters for the number of epochs (400, 600, 1000), the optimizer (RMSprop, Adam) and the number of nodes (20, 30). The optimized hyperparameters were 600, RMSprop and 20 for the number of epochs, optimizer and number of nodes, respectively. The accuracy of the ANN models were evaluated using coefficients of determination (R^2), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) on the validation data for each model (Eqs. 2–6). The trained and validated models were deployed on a globally accessible AISocket platform. The steam-assisted and microwave-assisted models can be located at https://www.aisocket.org/dashboard/005/ and https://www.aisocket.org/dashboard/007/, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(d_i - O_i)^2}{\sum_{i=1}^{n}(O_i)^2} \tag{2}$$

$$MAE = \frac{1}{n} \quad \sum\nolimits_{i=1}^{n} |d_i - O_i| \tag{3}$$

$$MAPE = \frac{100}{n} \quad \sum\nolimits_{i=1}^{n} \frac{O_i - d_i}{d_i} \tag{4}$$

$$MSE = \frac{\sum_{i=1}^{n}(d_i - O_i)^2}{n} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(d_i - O_i)^2}{n}} \tag{6}$$

where d_i and o_i are the observed and predicted outputs, and n is the

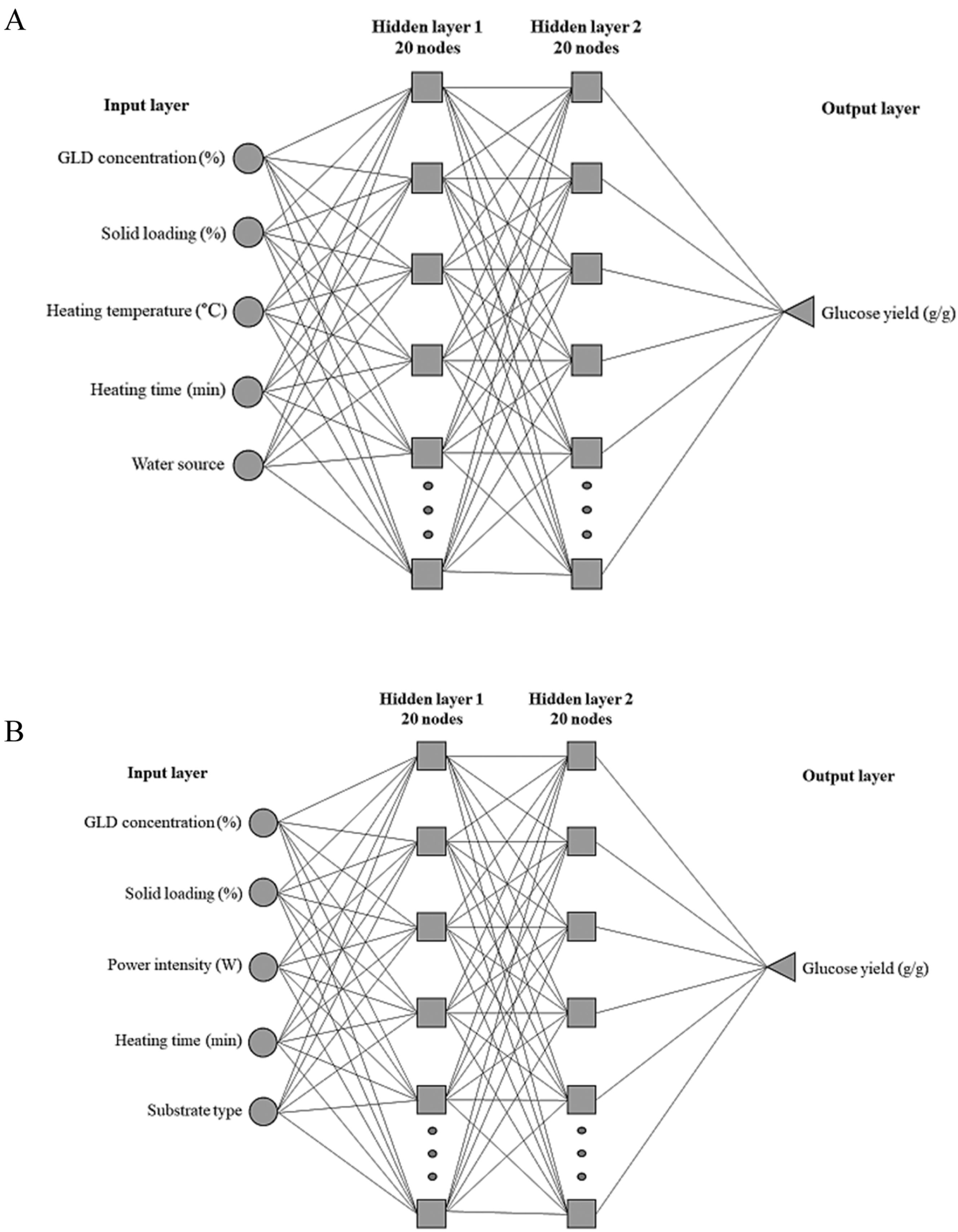

Fig. 1. Topology of Neural Networks used for the (A) steam- and (B) microwave-assisted Kraft waste-based model consisting of one input layer (five neurons), two hidden layers (20 nodes each) and one output layer (one neuron).

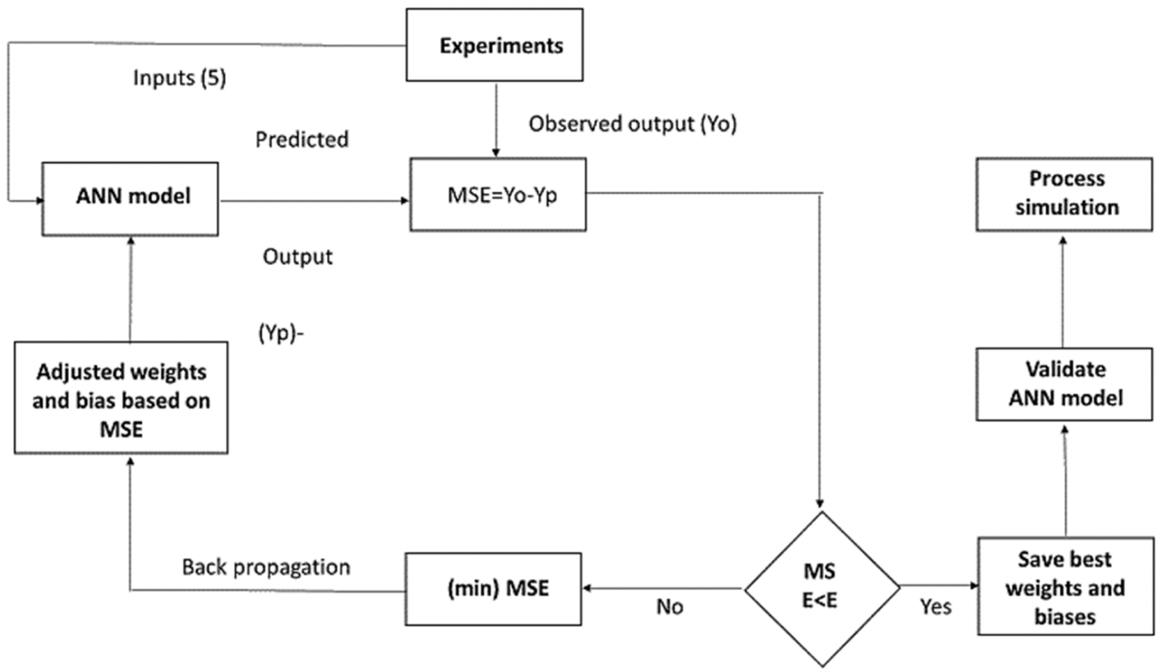

Fig. 2. The back propagation training flowchart for artificial neural network for the steam- and microwave-assisted Kraft waste-based model.

validation data set (20% of the data) for the ANN model.

2.4. Sensitivity analysis

Sensitivity analysis was utilized to determine the impact of fractional changes of the input parameters on the glucose response (Rorke et al., 2017). The variation of each input parameter ranged between its minimum (−1) and maximum (1) value, while maintaining all other inputs at their midpoint values (0). To explicate the functional relationships between pretreatment inputs and the glucose yield outputs, mathematical equations were derived from each model using curve fitting.

2.5. Analysis of data trends using large language model

The large language model, ChatGPT (version GPT-3.5 as of 24 May 2023, developed by OpenAI) was exposed to the datasets consisting of the fractional changes in the selected input parameters and its corresponding glucose yield outputs for both steam-assisted and microwave-assisted Kraft waste-based models of the present study. Search prompts were meticulously formulated and directed to the model in order to analyse the data trends of the glucose yield outputs in relation to its changes in input parameters. This approach enables interpretation of the relationships and patterns within the data through the perspective of the advanced LLM.

3. Results and discussion

3.1. Evaluation of developed ANN models

The developed ANN models were validated using the remaining 20% data sets not incorporated into the model training process. This validation focused on the prediction of glucose yield generation. The steam-based and microwave-assisted Kraft waste-based models yielded coefficients of determination (R^2) of 0.95 and 0.97, respectively as shown in Fig. 3A and B. This demonstrates the model's ability to effectively explain 95% and 97% of the variability observed in the data for the steam-based and microwave-based models, respectively. Typically, R^2 values falling within the range of 0.7–1 indicates a good model fit, as the R^2 value is a statistical measure of how well the regression line of the predicted data aligns with the actual data points. Furthermore, an additional metric analysis was conducted using the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) to assess the adequacy of the model's predictive performance. For the steam- and microwave-assisted Kraft waste-based models, the MAE suggests that, on average, the model's predictions deviate by approximately 0.065 and 0.043 units, respectively from the actual values. Lower MAE values indicate higher predictive accuracy, and the MAE values in the present study are relatively low, indicating a reasonably accurate model. The MAPE expresses the prediction errors as a percentage of the actual values. Our results indicate that the model's predictions exhibit a relative error of about 25.25% and 17.18% for the steam- and microwave-assisted models, respectively, when compared to the actual values. This suggests a moderate level of accuracy. The low MSE values of 0.007 for the steam-assisted model and 0.003 for the microwave-assisted model imply a good model fit. A low RMSE suggests that the model's predictions deviate by approximately 0.082 and 0.056 units from the actual values for the steam- and microwave-assisted models respectively. Both models demonstrate a reasonably accurate predictive performance, as evidenced by the low MAE, MSE and RMSE values, indicating minimal average prediction errors and a good model fit. The slight variation between experimental and predicted results may be attributed to the inherent complexities of a dynamic biological system that is characterized by constant change in the enzymatic saccharification process over time. This is dependent on the enzyme-substrate interactions, active energy exchanges and physicochemical responses. For this reason, the ANN modelling approach may not completely capture the intricacies of the biological system, leading to slight divergences between predicted and observed output. Nevertheless, the statistical indices suggest that the developed models displayed a strong correlation between the observed and predicted glucose outputs. As evidenced in both figures (Fig. 3A and B), a large majority of the data points are congregated along the predictive trend line, illustrating high accuracy for predicting the glucose yield when subjected to new process conditions. The robust relationship between the predicted and observed yields of the developed models can facilitate an appropriate navigation space needed for the pretreatment design studies. As such, the ANN predictive system aims to implement an initial preliminary screening tool for optimizing process

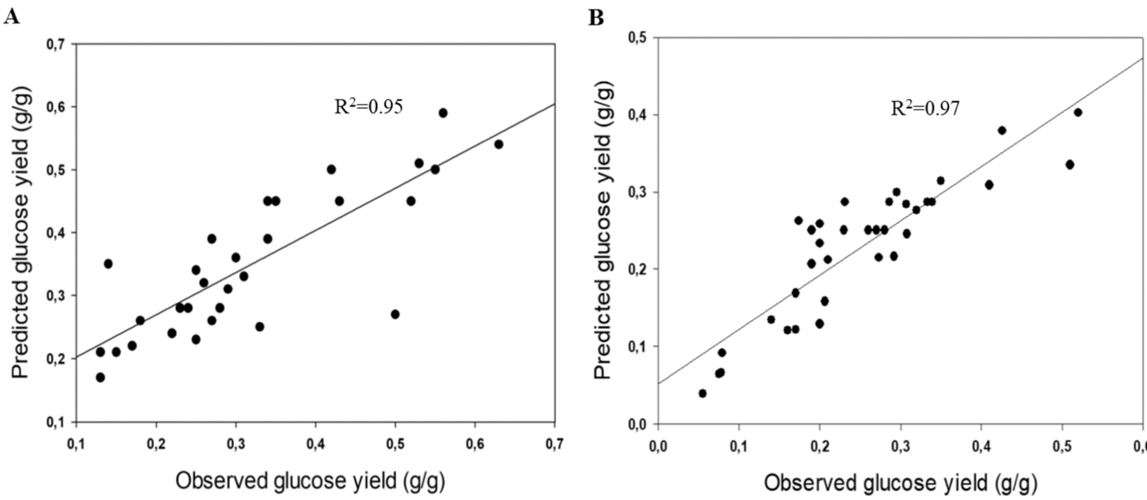

Fig. 3. Regression plots showing observed versus predicted glucose yield (g/g) for the (A) steam- assisted Kraft waste-based model and (B) microwave-assisted Kraft waste-based model. *Note*: The diagonal line depicts expectations under a one-to-one relationship for the predicted and observed yields.

parameters within lignocellulosic pretreatment technologies. These models have the potential to enhance resource efficiency, improve productivity and decrease process development costs. Kumari et al. (2019) optimized the pretreatment process for Leucaena leucocephala wood using three parameter inputs being catalyst concentration, duration and temperature with a total reducing sugar yield as its single response outcome. The results of these experiments were analyzed by three algorithms of neural networks such as Bayesian Regularization Neural Network (BRNN), Scaled Conjugate Gradient Neural Network (SCGNN), and Levenberg Marquardt Neural Network (LMNN) in which BRNN gave most accurate predictions for total reducing sugar yield with the lowest Root Mean Square Error (RMSE= 7.17 $\times 10^{-7}$) and Standard Error of Prediction percentage (SEP=3.84 $\times 10^{-7}$%). In a similar manner, Nikzad et al. (2015) reported on Response surface methodology (RSM) and ANN modelling of alkali pretreatment of rice husks for enhanced glucose and xylose yields. Both modeling methods were statistically compared by means of the R^2 and RSME. The RSM model gave R^2 (>0.83) and RMSE (>1.96) while the ANN model reported R^2 (>0.92) and RMSE (>1.36) value for the glucose and xylose yield prediction (Nikzad et al., 2015). It was therefore concluded that the ANN model observed a slightly higher prediction performance compared to RSM. Furthermore, Valim et al. (2017) reported on the use of ANN as prediction and fault detection tools for the delignification process of sugarcane bagasse subjected to alkaline H_2O_2. The performance of the ANN model was evaluated by the R^2 (0.99), MSE (0.004) and sum of squared errors (SSE=0.354). The values obtained for R^2 (close to 1) and the error indices (close to 0) indicated a good agreement of the theoretical and actual data.

3.2. Effect of changes in the input parameters on glucose yield

Sensitivity analysis was employed to determine the impact of the process input variables on the glucose yield and was carried out on the steam- (Fig. 4A-D) and microwave- (Fig. 4E-H) assisted Kraft waste-based pretreatment models. A sensitivity indicator represents the rate of change of the output, attributable to the fluctuations in the input parameters. An increased sensitivity to a parameter suggests that a slight variation in the input value can cause a significant change in the process output (Oh et al., 2003). On the other hand, a low sensitivity implies that even in the presence of large variations in the input parameter, a marginal change in the glucose output may occur. The functional relationships between the process inputs used within range and glucose output were derived from curve fitting for the steam- and microwave-assisted Kraft waste-based models (Table 2, Table 3).

For the steam-assisted Kraft waste-based model, the sensitivity analysis revealed that an increase in the GLD concentration from 1% to 30% resulted in a linear increase in the glucose yield from 0.146 g/g to 0.489 g/g and 0.125 g/g to 0.481 g/g when using PWW and pure water, respectively (Fig. 4A). This translates to a 3.3-fold and 3.8-fold increase from the baseline GLD concentration (1%) with the use of PWW and pure water, respectively. The exponential rise in glucose yield illustrates that it is largely dependent on the GLD concentration with a high sensitivity within this region. The high glucose yields obtained under the increasing GLD concentrations may be ascribed to the strong alkaline interactions of the GLD that cleaves specific bonds within the corn cob structure. This results in the disintegration of the lignocellulosic matrix and subsequent physical damage that facilitates easy penetration of the hydrolytic enzymes for improved sugar recovery (Kim et al., 2015). A further increase in the GLD concentration initiated a plateau in the glucose yield, followed by a slight decrease from 0.493 g/g to 0.481 g/g when the GLD concentration reached 40% (Fig. 4A). Given the primarily insoluble nature of GLD, its increase in concentration increases the viscosity by holding water tightly within the pretreatment slurry. This reduces the mass transfer capacity and hinders the diffusion of the GLD chemical to the CCW substrate. Furthermore, the temperature was maintained at its median value of 100 °C which may not be adequate for solubilization of the CCW substrate since heat transfer is impeded as a result of the high GLD concentrations (>30%). In a related study on alkaline pretreatment, Qing et al. (2016) observed an upward trend in the sugar yield when the Na_3PO_4 concentration was increased from 1% to 9%. A further increase in the Na_3PO_4 concentration above 9% resulted in a plateaued sugar yield (Qing et al., 2016). The relationship between the GLD concentration and glucose yield fit a rational model type for both the PWW and freshwater pretreatments (Table 2).

Conversely, a hyperbolic decline relationship (Table 2) was observed in Fig. 4B where an increase in the solid loading from 2% to 20% triggered a decrease in glucose yield from 0.537 g/g to 0.379 g/g and 0.522 g/g to 0.361 g/g when subjected to PWW and fresh water, respectively. The low SL (2%) facilitates a pretreatment medium that is relatively fluid in nature which ensures optimal GLD-substrate interactions by maximizing mass and heat transfer. As the SL increases (>2%), the liquid volume is reduced due to absorption, culminating in a viscous suspension with low diffusion capacity. This indicates that a SL < 2% enables a maximum glucose yield, resulting in a reduced resource utilization for improved process economics. Notably, Raghavi et al. (2016) showed that an incline in the NaOH pretreated sugarcane waste from 10% to 15% SL decreased the sugar yield from 0.70 g/g to< 0.50 g/g.

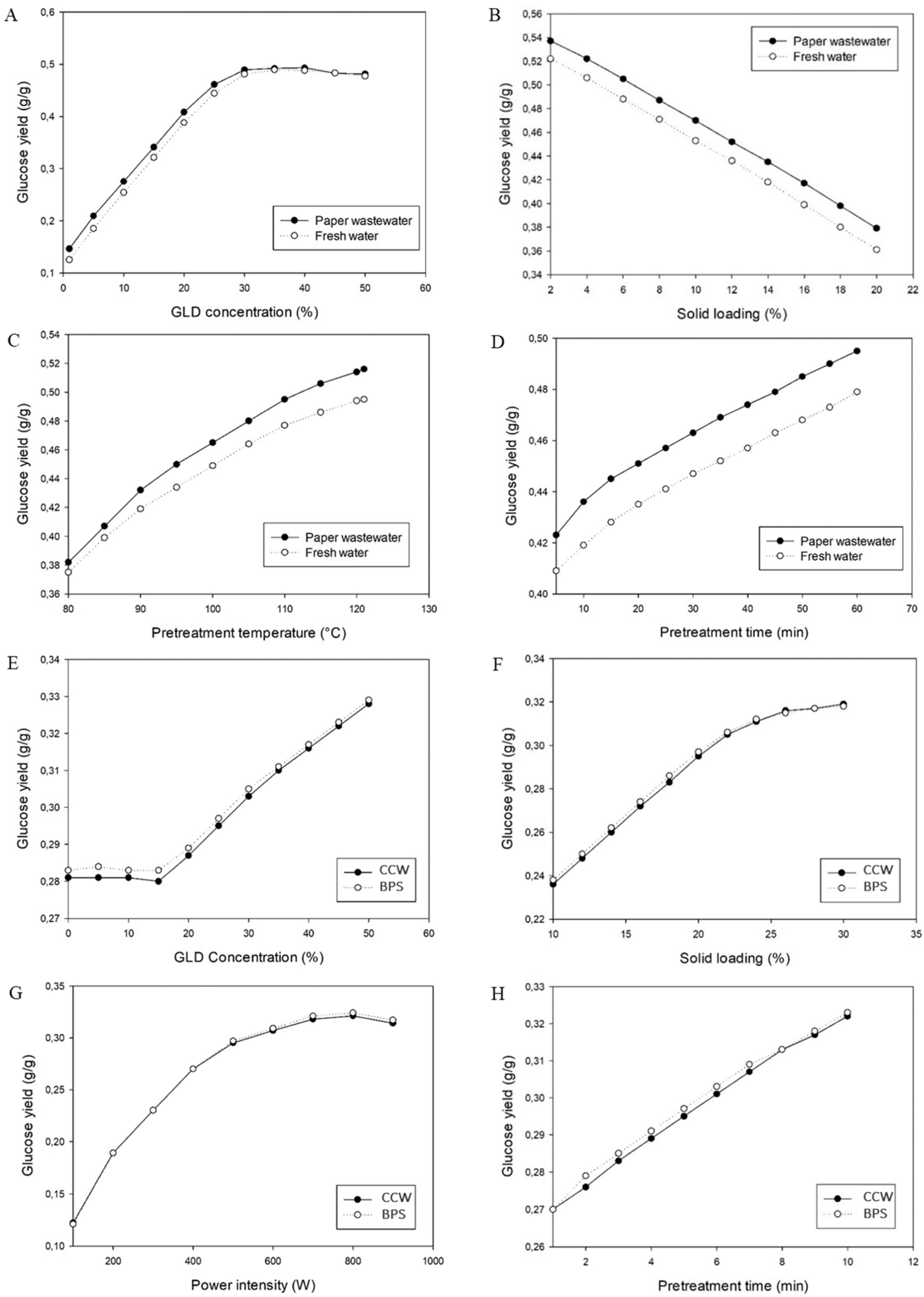

Fig. 4. Effect of fractional changes of input parameters from the steam-assisted (A-D) and microwave-assisted Kraft waste-based pretreatments (E-H) on the glucose yield.

Table 2
Steam-assisted Kraft waste-based model equations for the effect of fractional changes in the input parameters on the process output.

Water source	Input	Model equation	Model type	Fitted model	R^2
PWW	(a) GLD concentration	$y = \frac{a+bx}{1+cx+dx^2}$	Rational	$y = \frac{0.13+0.01x}{1-0.02x+0.001x^2}$	0.99
	(b) Solid loading	$y = q_0\left(1+\frac{bx}{a}\right)^{\left(-\frac{1}{b}\right)}$	Hyperbolic decline	$y = 0.55\left(1-\frac{1.44x}{68.49}\right)^{\left(\frac{1}{1.44}\right)}$	0.99
	(c) Temperature	$y = \frac{\theta x^\eta}{\kappa^\eta + x^\eta}$	DR-Hill-Zerbackground	$y = \frac{0.59x^{3.22}}{66.27^{3.22}+x^{3.22}}$	0.99
	(d) Pretreatment time	$y = \frac{a+bx}{1+cx+dx^2}$	Rational	$y = \frac{0.39+0.11x}{1+0.24x-0.0004x^2}$	0.99
Fresh water	(a) GLD concentration	$y = \frac{a+bx}{1+cx+dx^2}$	Rational	$y = \frac{0.11+0.01x}{1-0.02x+0.001x^2}$	0.99
	(b) Solid loading	$y = q_0\left(1+\frac{bx}{a}\right)^{\left(-\frac{1}{b}\right)}$	Hyperbolic decline	$y = 0.54\left(1-\frac{1.42x}{65.46}\right)^{\left(\frac{1}{1.44}\right)}$	0.99
	(c) Temperature	$y = \frac{\theta x^\eta}{\kappa^\eta + x^\eta}$	DR-Hill-Zerbackground	$y = \frac{0.57x^{2.93}}{64.39^{2.93}+x^{2.93}}$	0.99
	(d) Pretreatment time	$y = a + br^x + cx$	Exponential plus linear	$y = 0.42 - (0,02)(0,89)^x + 0.001x$	0.99

Footnote: PWW = paper wastewater, GLD= green liquor dregs, R^2 = Coefficient of determination.

Table 3
Microwave-assisted Kraft waste-based model equations for the effect of fractional changes in the input parameters on the process output.

Substrate type	Input	Model equation	Model type	Fitted model	R^2
CCW	(a) Time	$y = \frac{a+bx}{1+cx+dx^2}$	Rational	$y = \frac{2.27-17.18x}{1-60.29x+0.77x^2}$	0.99
	(b) Solid loading	$y = \frac{a}{(1+e^{b-cx})^{\frac{1}{d}}}$	Richards	$y = \frac{0.32}{(1+e^{9.46-0.42x})^{\frac{1}{17.56}}}$	0.99
	(c) Power intensity	$y = \frac{1}{A+B\ln(x)+C(\ln(x))^3}$	Steinhart-Hart equation	$y = \frac{1}{41.63-8.66\ln(x)+6.5(\ln(x))^3}$	0.99
	(d) GLD concentration	$y = \alpha + \frac{\theta x^\eta}{\kappa^\eta + x^\eta}$	DR-Hill	$y = 0.28 + \frac{0.06x^{3.65}}{34.65^{3.65}+x^{3.65}}$	0.99
BPS	(a) Time	$y = \gamma + (1-\gamma)\Gamma(\alpha, \beta x)$	DR-Gamma	$y = 0.26 + (1-0.26)x0.26(0.74, 0.003x)$	0.99
	(b) Solid loading	$y = \frac{a}{(1+e^{b-cx})^{\frac{1}{d}}}$	Richards	$y = \frac{0.32}{(1+e^{9.19-0.42x})^{\frac{1}{17.30}}}$	0.99
	(c) Power intensity	$y = a + br^x + cx$	Exponential plus linear	$y = 0.89 - (0,85)(1)^x - 0.0004x$	0.99
	(d) GLD concentration	$y = \alpha + \frac{\theta x^\eta}{\kappa^\eta + x^\eta}$	DR-Hill	$y = 0.28 + \frac{0.06x^{3.54}}{35.26^{3.54}+x^{3.54}}$	0.99

Footnote: CCW = Corn cob waste, BPS= Banana pseudostem, GLD= green liquor dregs, R^2 = Coefficient of determination.

On the other hand, a steady rise in the glucose yield from 0.382 g/g to 0.516 g/g (PWW) and 0.375 g/g to 0.495 g/g (fresh water) was shown when the pretreatment temperature was increased beyond 80 °C till 121 °C (Fig. 4C). This interaction between the temperature and glucose yield was illustrated by a DR-Hill-Zer background relationship (Table 2). Exposure to high temperature induces rapid reaction rates and enhanced solubilization, attributable to the change in chemical properties of water by ionization (Toscan et al., 2019).

Likewise, when the pretreatment time increased from 5 min to 60 min, the glucose yield increased linearly from 0.423 g/g to 0.495 g/g and 0.409 g/g to 0.479 g/g when using PWW and fresh water, respectively (Fig. 4D). An extended pretreatment time increases the contact time for the GLD, fostering a positive influence on the fractionation of the LCB. The rational model best illustrated the relationship between the pretreatment time and glucose yield with PWW as a water source (Table 2). On the other hand, the exponential plus linear model type expressed the relationship between the pretreatment time and glucose yield when fresh water was utilized in the pretreatment system (Table 2).

With regards to the sensitivity analysis for the microwave-assisted Kraft waste-based model, similar glucose yields of 0.281 g/g and 0.283 g/g from CCW and BPS, respectively, were depicted when the GLD concentration increased from 0% to 15% (Fig. 4E). The GLD concentration to SL ratio translates to < 1:1, suggesting that the pretreatment process may be limited by the low GLD concentrations and possibly permeates only a fraction of the total median biomass (20%) within solution. Thereafter, a sharp linear increase was observed in the glucose yield of the CCW (0.287 g/g to 0.328 g/g) and BPS (0.289 g/g to 0.329 g/g) when the GLD concentration was further elevated from 20% till 50% (Fig. 4E). Interestingly, when the GLD to SL ratio was initiated at a > 1:1 ratio, the glucose yield showed a strong positive correlation. Accordingly, the DR-Hill model was used to describe the relationship between the GLD concentration and glucose yield using both the CCW and BPS as substrate (Table 3).

In another instance, an incline in the solid loading from 10% to 24% initiated a linear increase in the glucose yield from 0.236 g/g to 0.311 g/g and 0.238 g/g to 0.312 g/g after pretreatment of CCW and BPS, respectively (Fig. 4F). The GLD to SL ratio when the SL range was between 10% and 24% was denoted at > 1:1, resulting in the induction of saturated GLD solutions for optimal lignocellulosic disbanding with significant cellulose release. A further increase in the SL from 24% to 30% caused the glucose release from CWW (0.311 g/g to 0.319 g/g) and BPS (0.312 g/g to 0.318 g/g) to increase slightly thereafter. Despite a higher SL (> 24%), the pretreatment threshold may have been reached due to GLD being the limiting factor. This interaction was depicted using the Richards model for both CCW and BPS (Table 3).

Remarkably, exponential increases in the glucose yields from 0.122 g/g to 0.318 g/g (CCW) and 0.121 g/g to 0.321 g/g (BPS) were shown when the power intensity was increased from 100 W to700 W (Fig. 4G). This translates to a 2.6-fold and 2.7-fold increase from the baseline power intensity (100 W) when using CCW and BPS, respectively. Microwave irradiation promotes the breakdown of the lignocellulosic constituents by the introduction of molecular collisions caused by dielectric polarization on the chemical covalent bonding of the feedstock (Sorn et al., 2019). A further increase in the power intensity from 700 W to 800 W resulted in a slight plateau of the glucose yield where CCW only released 0.318 g/g to 0.321 g/g and BPS produced 0.321 g/g to 0.324 g/g. Interestingly, an increase beyond 800 W reduced the glucose yields for both the CCW (0.314 g/g) and BPS (0.317 g/g). In the same vein, Sewsynker-Sukai and Gueguim Kana (2018) depicted that an incline in power intensity beyond 700 W did not significantly influence the sugar yield. This was ascribed to the high microwave irradiation that accelerates the evaporation of water and adversely affects the homogeneity of the pretreatment slurry and penetration of the microwaves to the biomass. This correlation was best described by a Steinhart-Hart equation and exponential plus linear model types for the CCW and BPS substrates, respectively (Table 3).

Moreover, when the pretreatment time was adjusted from 1 min to 10 min, the glucose obtained from the CCW drastically increased from 0.27 g/g to 0.322 g/g while the BPS gave 0.27 g/g to 0.323 g/g (Fig. 4H). Lin et al. (2015) demonstrated a similar effect of pretreatment time on water hyacinth using microwave pretreatment and revealed that increments from 5 to 10 min resulted in a 51.85% increase in the reducing sugar yield. However, a further extension in the pretreatment time did not significantly impact the sugar yield and may be attributed to the degradation of monosaccharides into other by-products such as acetic acid and furfural. As a result, the rational model was used to illustrate the interaction between the pretreatment time and glucose yield using CCW while the DR-gamma model fit the association with BPS (Table 3).

3.3. *Critical assessment of large language model interpretations: comparisons and insights in lignocellulosic pretreatment research*

This study utilized a LLM to extract knowledge on the impact of several physicochemical input parameters on fermentable sugar yields during lignocellulosic pretreatment, specifically focusing on the complete Kraft waste-based lignocellulosic pretreatment. The model's responses were based on meticulously designed input prompts. Subsequently, the prompts were compared to the authors findings detailed in *Section 3.2* , drawn from domain expertise and direct analysis of study data. A selection of significant input parameters impacting Kraft waste-based pretreatment of CCW for improved glucose recovery, as interpreted by the LLM, are discussed below:

For instance, when prompted on the data trends with regards to the effect of varied solid loading during steam-assisted Kraft waste-based pretreatment on glucose yield, the LLM elucidated that the decrease in glucose yield with increasing solid loading can be attributed to several factors. One of the notable factors is the decrease in mass and heat transfer efficiency. As the solid loading increases, the liquid volume decreases due to absorption, which in turn results in a more viscous suspension with low diffusion capacity (See Supplementary material). The model interpretation aligns with previous studies underscoring that increased solid loading results in reduced substrate-pretreatment chemical contact and lower glucose yields (Raghavi et al., 2016; Sewsynker-Sukai and Gueguim Kana, 2017b). Moreover, the model provided additional insight, stating that high solid loadings result in a higher degree of substrate structural heterogeneity, which can potentially inhibit enzymatic hydrolysis, a known occurrence in LCB pretreatment (Jeoh et al., 2017). Similarly, ChatGPT interpreted the linear increase in glucose yield with increasing GLD concentration within the steam-assisted Kraft waste-based pretreatment. The enhanced delignification and biomass fractionation at higher GLD concentrations have been attributed to increased alkaline conditions, which facilitate the breakdown of lignin and hemicellulose, making cellulose more accessible for enzymatic hydrolysis. However, excessive GLD concentration can lead to increased slurry viscosity and reduced mass transfer, ultimately decreasing the overall efficiency of the pretreatment process (See supplementary material). The LLM highlights the importance of optimizing GLD concentration during pretreatment to achieve maximum glucose yields, while mitigating the negative effects associated with excessive GLD concentrations.

In the same vein, the model interpreted the effects of the fractional changes in the power intensity and its influence on the glucose yield during microwave-assisted Kraft waste-based pretreatment. It detailed that an increasing microwave power intensity can enhance biomass pretreatment effectiveness due to higher energy input, leading to improved sugar yields. The model further indicated that beyond an optimal power intensity, improvements in the glucose yield may be insignificant and could possibly lead to decreased yields (See supplementary material). In this context, the ChatGPT responses coincide with previous studies on the influence of microwave power intensity on the pretreatment of LCB (Sewsynker-Sukai and Gueguim Kana, 2018; Laltha et al., 2021).

The LLM knowledge generated from the search prompts were evaluated in light of the authors interpretations of the findings in this study (*Section 3.2*). When comparing the LLM responses to the contextual interpretation, significant parallels were observed. The model was able to effectively recognize and interpret the data trends presented for the parameter impact on the glucose yields of the steam and microwave-assisted Kraft waste-based pretreatment. This approach also provided a deeper understanding of the underlying mechanisms in the pretreatment process for improved predictive modelling in future research. Therefore, the LLM contributes substantially to the existing body of knowledge on lignocellulosic pretreatment available within the public domain.

Despite these remarkable insights, it is important to note that LLMs present certain limitations due to its reliance on the statistical associations with existing literature rather than depth of understanding and contextual awareness, possibly leading to misinterpretation of data (Ray, 2023). Furthermore, it lacks real-time information since its training is reliant on a static dataset that has a knowledge cut-off date, specifically the year 2021. For this reason, expert human knowledge inputs are necessary to augment LLM interpretations. This synergistic approach underscores the use of predictive LLM modelling technologies to identify patterns, extract key insights and provide broader context, while human experts use critical thinking and domain knowledge to validate and verify its sources to ensure factual accuracy and reliability.

4. Conclusion

The developed models demonstrated R^2 values > 0.95 for the predicted and observed glucose outputs. Sensitivity analysis exhibited increased response to GLD concentration and power intensity for the steam and microwave models, respectively. Furthermore, ChatGPT presented key insights that aligned with the study's contextual interpretations, substantiating its value as a complementary analytical tool. By harnessing the objectivity of virtual tools, valuable knowledge on behavioural patterns and outcomes associated with waste-based lignocellulosic pretreatment prior to process development, optimization and scale-up will be generated. This facilitates informed decision making, pretreatment design improvement, automated assessments of economic outlooks and productivity within lignocellulosic biorefineries.

CRediT authorship contribution statement

Anthea N. David: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration, Funding acquisition. **Y. Sewsynker-Sukai:** Validation, Data curation, Writing - review & editing, Supervision, Project administration. **E.L. Meyer:** Writing - review & editing. **E.B. Gueguim Kana:** Validation, Data curation, Writing - review & editing, Resources, Supervision, Project administration.

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

Data will be made available on request.

Acknowledgments

This work is based on research supported in part by the National Research Foundation (NRF) of South Africa (Grant number 122080). Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.indcrop.2023.117686.

References

Chang, C.W., Yu, W.C., Chen, W.J., Chang, R.F., Kao, W.S., 2011. A study on the enzymatic hydrolysis of steam exploded napiergrass with alkaline treatment using artificial neural networks and regression analysis. J. Taiwan Inst. Chem. Eng. 42 (6), 889–894.
Chen, S., Ren, Y., Friedrich, D., Yu, Z., Yu, J., 2020. Sensitivity analysis to reduce duplicated features in ANN training for district heat demand prediction. Energy AI 2, 100028.
David, A.N., Sewsynker-Sukai, Y., Sithole, B., Gueguim Kana, E.B., 2020. Development of a green liquor dregs pretreatment for enhanced glucose recovery from corn cobs and kinetic assessment on various bioethanol fermentation types. Fuel 274, 117797.
David, A.N., Sewsynker-Sukai, Y., Gueguim Kana, E.B., 2021. Development of Kraft waste-based pretreatment strategies for enhanced sugar recovery from lignocellulosic waste. Ind. Crop. Prod. 174, 114222.
Jeoh, T., Cardona, M.J., Karuna, N., Mudinoor, A.R., Nill, J., 2017. Mechanistic kinetic models of enzymatic cellulose hydrolysis-a review. Biotechnol. Bioeng. 114 (7), 1369–1385.
Kim, J.S., Lee, Y.Y., Kim, T.H., 2015. A review on alkaline pretreatment technology for bioconversion of lignocellulosic biomass. Bioresour. Technol. 199, 42–48.
Kumari, N., Garg, S., Singhal, A., Kumar, M., Bhattacharya, M., Jha, P.K., Devendra, K.C., Thakur, I.S., 2019. Optimizing pretreatment of Leucaena leucocephala using artificial neural networks (ANNs). Bioresour. Technol. Rep. 7, 100289.
Laltha, M., Sewsynker-Sukai, Y., E.B, Gueguim Kana, 2021. Development of microwave-assisted alkaline pretreatment methods for enhanced sugar recovery from bamboo and corn cobs: process optimization, chemical recyclability and kinetics of bioethanol production. Ind. Crops Prod. 174, 114166.
Laltha, M., Sewsynker-Sukai, Y., Gueguim Kana, E.B., 2022. Innovative microwave-assisted iodized table salt or paper wastewater pretreatments for enhanced sugar recovery from banana pseudostem. Biomass-.-. Convers. Biorefin. 1–15.
Lee, K.M., Zanil, M.F., Chan, K.K., Chin, Z.P., Liu, Y.C., Lim, S., 2020. Synergistic ultrasound-assisted organosolv pretreatment of oil palm empty fruit bunches for enhanced enzymatic saccharification: An optimization study using artificial neural networks. Biomass-.-. Bioenerg. 139, 105621.
Lin, R., Cheng, J., Song, W., Ding, L., Xie, B., Zhou, J., Cen, K., 2015. Characterisation of water hyacinth with microwave-heated alkali pretreatment for enhanced enzymatic digestibility and hydrogen/methane fermentation. Bioresour. Technol. 182, 1–7.
McKendry, P., 2002. Energy production from biomass (part 1): overview of biomass. Bioresour. Technol. 83 (1), 37–46.
Moodley, P., Rorke, D.C., Gueguim Kana, E.B., 2019. Development of artificial neural network tools for predicting sugar yields from inorganic salt-based pretreatment of lignocellulosic biomass. Bioresour. Technol. 273, 682–686.
Muthuvelu, K.S., Rajarathinam, R., Kanagaraj, L.P., Ranganathan, R.V., Dhanasekaran, K., Manickam, N.K., 2019. Evaluation and characterization of novel sources of sustainable lignocellulosic residues for bioethanol production using ultrasound-assisted alkaline pretreatment. Waste Manag 87, 368–374.
Nikzad, M., Movagharnejad, K., Talebnia, F., Aghaiy, Z., Mighani, M., 2015. Modeling of alkali pretreatment of rice husk using response surface methodology and artificial neural network. Chem. EnG. Commun. 202 (6), 728–738.
Nunes, L.J.R., Causer, T.P., Ciolkosz, D., 2020. Biomass for energy: a review on supply chain management models. Renew. Sust. Energy Rev. 120, 109658.
Oh, S.E., Van Ginkel, S., Logan, B.E., 2003. The relative effectiveness of pH control and heat treatment for enhancing biohydrogen gas production. Environ. Sci. Technol. 37 (22), 5186–5190.
Patel, A., Shah, A.R., 2021. Integrated lignocellulosic biorefinery: gateway for production of second generation ethanol and value added products. J. Bioresour. Bioprod. 6 (2), 108–128.
Qing, Q., Zhou, L., Huang, M., Guo, Q., He, Y., Wang, L., Zhang, Y., 2016. Improving enzymatic saccharification of bamboo shoot shell by alkalic salt pretreatment with H_2O_2. Bioresour. Technol. 201, 230–236.
Raghavi, S., Sindhu, R., Binod, P., Gnansounou, E., Pandey, A., 2016. Development of a novel sequential pretreatment strategy for the production of bioethanol from sugarcane trash. Bioresour. Technol. 199, 202–210.
Ray, P.P., 2023. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber-Phys. Syst. 3, 121–154.
Rebello, S., Anoopkumar, A.N., Aneesh, E.M., Sindhu, R., Binod, P., Pandey, A., 2020. Sustainability and life cycle assessments of lignocellulosic and algal pretreatments. Bioresour. Technol. 301, 122678.
Rego, A.S., Valim, I.C., Vieira, A.A., Vilani, C., Santos, B.F., 2018. Optimization of sugarcane bagasse pretreatment using alkaline hydrogen peroxide through ANN and ANFIS modelling. Bioresour. Technol. 267, 634–641.
Rorke, D.C., Suinyuy, T.N., Gueguim Kana, E.B., 2017. Microwave-assisted chemical pre-treatment of waste sorghum leaves: process optimization and development of an intelligent model for determination of volatile compound fractions. Bioresour. Technol. 224, 590–600.
Sebayang, A.H., Masjuki, H.H., Ong, H.C., Dharma, S., Silitonga, A.S., Kusumo, F., Milano, J., 2017. Optimization of bioethanol production from sorghum grains using artificial neural networks integrated with ant colony. Ind. Crop. Prod. 97, 146–155.
Sewsynker-Sukai, Y., Gueguim Kana, E.B., 2017a. Does the volume matter in bioprocess model development? An insight into modelling and optimization of biohydrogen production. Int. J. Hydrog. Energ. 42 (9), 5780–5792.
Sewsynker-Sukai, Y., Gueguim Kana, E.B., 2017b. Optimization of a novel sequential alkalic and metal salt pretreatment for enhanced delignification and enzymatic saccharification of corn cobs. Bioresour. Technol. 243, 785–792.
Sewsynker-Sukai, Y., Gueguim Kana, E.B., 2018. Microwave-assisted alkalic salt pretreatment of corn cob wastes: Process optimization for improved sugar recovery. Ind. Crop Prod. 125, 284–292.
Sindhu, R., Binod, P., Pandey, A., 2016. A novel sono-assisted acid pretreatment of chilli post harvest residue for bioethanol production. Bioresour. Technol. 213, 58–63.
Sorn, V., Chang, K.L., Phitsuwan, P., Ratanakhanokchai, K., Dong, C.D., 2019. Effect of microwave-assisted ionic liquid/acidic ionic liquid pretreatment on the morphology, structure, and enhanced delignification of rice straw. Bioresour. Technol. 293, 121929.
Toscan, A., Fontana, R.C., Andreaus, J., Camassola, M., Lukasik, R.M., Dillon, A.J.P., 2019. New two-stage pretreatment for the fractionation of lignocellulosic components using hydrothermal pretreatment followed by imidazole delignification: focus on the polysaccharide valorization. Bioresour. Technol. 285, 121346.
Valim, I.C., Fidalgo, J.L., Rego, A.S., Vilani, C., Martins, A.R.F., Santos, B.F., 2017. Neural network modeling to support an experimental study of the delignification process of sugarcane bagasse after alkaline hydrogen peroxide pre-treatment. Bioresour. Technol. 243, 760–770.
Vani, S., Sukumaran, R.K., Savithri, S., 2015. Prediction of sugar yields during hydrolysis of lignocellulosic biomass using artificial neural network modeling. Bioresour. Technol. 188, 128–135.