ORIGINAL ARTICLE

Optimizing Fermentation Parameters for Bioethanol Production from Areca Nut Leaves using Artificial Neural Networks and Response Surface Methodologies

Amal Bahafi[1], Uday M. Muddapur[2], Veeranna S. Hombalimath[2], Ibrahim Ahmed Shaikh[3], S. Manjunath[2], Laxmikant R. Patil[2], Aejaz A. Khan[4], S. M. Shakeel Iqubal[4]

[1]Department of Pharmaceutical Chemistry, Ibn Sina National College for Medical Studies, Jeddah, Saudi Arabia, [2]Department of Biotechnology, KLE Technological University, BVB Campus, Hubli, Karnataka, India, [3]Department of Pharmacology, College of Pharmacy, Najran University, Najran, Saudi Arabia, [4]Department of General Science, Ibn Sina National College for Medical Studies, Jeddah, Saudi Arabia

Abstract

Aims: This study focused on the production of bioethanol from Areca nut leaves, a significant cultivated feedstock. The research covered the entire process, from collecting the Areca nut leaves to purifying the produced bioethanol. **Materials and Methods**: The Areca nut leaves were pre-treated with sulfuric acid and sodium hydroxide, followed by enzymatic hydrolysis using cellulose enzymes. The hydrolyzed biomass was then fermented by *Saccharomyces cerevisiae* for 12–72 h to produce bioethanol. The produced bioethanol was purified through distillation using a rotary flask evaporator. To optimize the fermentation process and bioethanol production, the researchers employed two modeling approaches: Artificial neural networks (ANN) and response surface methodology (RSM). Variables such as pH, fermentation time, and disodium hydrogen phosphate (Na_2HPO_4) concentration, identified from the Plackett-Burman design, were optimized using the central composite design of RSM. **Results and Discussion:** The R^2 value for the RSM model was 91.72%, and the adjusted R^2 was 84.72%. In addition, an ANN algorithm model with 3 input neurons, 10 hidden layer neurons, and 1 output neuron was developed to investigate the relationship between bioethanol production and fermentation parameters. The ANN model achieved an R^2 of 99.78%, indicating higher accuracy and reliability compared to the RSM approach. The optimal conditions for bioethanol production were identified as pH 5.5, 60 h fermentation time, and 0.45 g of Na_2HPO_4. Under these conditions, the experimental bioethanol concentration reached 36.54 g/L. **Conclusion:** This study demonstrates the effective utilization of Areca nut leaves, a readily available agricultural waste, to produce bioethanol. The combination of statistical and machine learning techniques, such as ANN and RSM, allowed for the optimization of the fermentation process and the enhancement of bioethanol yield, showcasing the potential of this approach for sustainable biofuel production.

Key words: Artificial neural networks, areca nut leaves, bioethanol, enzymatic hydrolysis, response surface methodology methodologies

INTRODUCTION

The detrimental environmental effects, notably global warming resulting from the overreliance on fossil fuels, have made scientists to seek alternative energy sources such as renewable energy, Bioethanol derived from various lignocellulosic biomasses including sugarcane bagasse, rice straw, corn straw, and wheat straw considered agro-industrial waste emerges as a promising sustainable substitute for fossil fuels.[1] Over the past five decades, numerous technologies have been developed to efficiently convert biomass into biofuels, aiming to make bioethanol a cost-competitive fuel in the contemporary fuel

Address for correspondence:
Veeranna S. Hombalimath, Department of Biotechnology, KLE Technological University, BVB Campus, Hubli - 580 031, Karnataka, India.
E-mail: hombalimath@kletech.ac.in; shakeeliqubal@gmail.com

Received: 07-08-2024
Revised: 23-10-2024
Accepted: 02-11-2024

market.[2] Lignocellulosic biomass comprises cellulose (40–60%), hemicellulose (20–40%), and lignin forming elastic structure.[3] Cellulose, a polymer of glucose glycosidic bonds, fosters extensive hydrogen bonding, resulting in compact crystalline cell biological degradation. Hemicellulose, predominantly composed of xylene (α-1,4 linkages) branches of mannose, arabinose, galactose, and glucuronic acid, exhibits degrees of branch on biomass origin. Lignin, a large aromatic and hydrophobic biopolymer, cross-links with fortifying the cell wall and imparting mechanical strength.[4] Areca nut was predominantly cultivated majority in Karnataka, Kerala, and Assam, India's areca nut production was approximately 83% of the total cultivation area. Nearly 4 lakh hectares under areca nut production of roughly 4.78 lakh tons from India, Karnataka leads in both cultivation areas and followed by Kerala and Assam.

Composition of lignocellulosic biomass

Lignocellulosic materials were primarily composed of three components: Cellulose, hemicellulose, and lignin. Together, cellulose and hemicellulose make up about 70% of the total biomass. These components are intricately bonded to lignin through covalent and hydrogen bonds, which enhance the material's structural integrity and resistance to treatment.

The adverse environmental impacts, particularly the exacerbation of global warming due to dependence on fossil fuels, have driven scientists to explore alternative energy sources, such as energy. Bioethanol, extracted from diverse lignocellulosic biomasses, such as sugarcane bagasse, corn straw, and wheat straw recognized as agro-industrial byproducts – emerges as a pro-sustainable alternative to fossil fuels.[1] Over the past 50 years, technologies has been developed to efficiently convert biomass into biofuels, with the objective of bioethanol as a competitive fuel option in the contemporary energy market.[5,6] Lignocellulosic biomass mainly consists of cellulose (40–60%), hemicellulose (20–40%), and lignin forming a robust structure.[3] Cellulose, a glucose polymer links glycoside bonds and facilitates extensive hydrogen bonding, leading to the formation of composite cellulose that resists biological degradation. Hemicellulose, primarily comprised of xylene with diverse branches of mannose, arabinose, galactose, and glucuronic acid, exhibits various branching depending on the biomass source. Lignin, a substantial aromatic and hydrophobia interacts with hemicellulose, reinforcing the cell wall and providing mechanical strength). Areca nut a crucial commercial crop in India dominates the global problems.

MATERIALS AND METHODS

The Areca nut leaves were gathered from Sirsi, Uttara Kannada, India, which were dried to remove moisture content was finely ground into particles sized 1–2 mm.

Enzymatic hydrolysis

Ten g of oven-dried Areca nut leaves were dissolved in 100 mL of a sodium acetate (CH_3COONa) solution containing 0.68 g of solid CH_3COONa. The pH was adjusted to a range of 4.0–6.0 using 1.0 M NaOH and 1.0 M H_2SO_4. Another 5 g of cellulose enzyme was added to the solution, and the flask was sealed with cotton foil. The mixture was incubated in a shaker at 37°C and 150 rpm for a specified duration. Samples were periodically withdrawn for glucose testing at regular intervals of 12–24 h.

PB design and central composite design (CCD)

Response surface methodology (RSM) involves a set of experimental techniques was used to assess the relationship between experimental factors and determine their responses. The significant variables influencing bioethanol production were screened using the Plackett-Burman design (PBD). This design was experimented using Minitab software employed for (95% confidence level). The acceptance criterion for the predicted model was based on an adjusted coefficient of regression (R^2adj) which was exceeding 0.95. Variables with $P = 0.05$ for PBD and 0.01 for the CCD were considered to have a significant effect on the response. The independent variables selected for this study included physical parameters, such as pH and fermentation time, as well as media components such as yeast extract, ammonium chloride, disodium hydrogen phosphate (Na_2HPO_4), and potassium dihydrogen phosphate. In addition, a central composite rotatable design with three independent variables at five levels each was conducted. This experimental setup was aimed to establish a second-degree

Table 1: The effect of enzymatic hydrolysis period on glucose content

Sample	Enzymatic hydrolysis period (h)	Glucose content (g/L)
Areca nut leaves	24	31.55
	48	47.45
	72	56.76

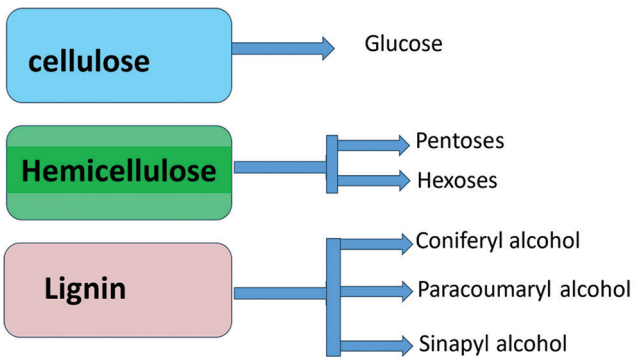

Figure 1: Lignocellulosic biomass composition (cellulose, hemicellulose and lignin)

polynomial model equation that describes bioethanol production as a function of three independent variables: pH, fermentation time, and Na_2HPO_4, for the fermentation process. All experiments were conducted randomly, and the resulting data were analyzed using Minitab software.

Artificial neural network (ANN)

ANNs were widely used for optimizing process parameters in fermentation processes intended at bioethanol production.[7-9] In this study, ANN represents an intelligence technique, commonly employed for modeling complex phenomena involving numerous process parameters.[10,11] The predictive capability of ANN relies on experimental data and subsequent validation with independent data.[12] ANN tool was used to address non-linear models by assessing relationships between input and output parameters, even when the data are intricate and incomplete patterns.[13-15]

Figure 2: Wet and dry arecanut leaves

RESULTS AND DISCUSSION

The pre-treatment of Areca nut leaves was carried by enzymatic hydrolysis and further used for the fermentation process. Enzymatic hydrolysis was carried out for 24 h, 48 h, and 72h. It was observed that the Areca nut leaves residues exhibited higher glucose yield, indicating an extensive reaction between the cellulose enzyme and the Areca nut leaves (Figures 1 and 2). The analysis of glucose content for different Areca nut samples was found to be 31.55 g/L, 47.45 g/L, and 56.76 g/L at 24 h, 48 h, and 72 h, respectively.

The results showed a continuous increase in glucose content with the extension of the hydrolysis period. The glucose content was doubled within a 48-h hydrolysis.

RSM modeling

The process of optimization of the media components for the highest bioethanol production was carried out by selecting significant process parameters, such as pH, fermentation time, and Na_2HPO_4.[16-18] The model given by the equation indicates bioethanol production as a function of pH, Na_2HPO_4, and fermentation time.

The statistical significance of the quadratic regression model was verified using analysis of variance and Fisher's test (F). A high F-value and a low *P*-value indicate that the model was statistically significant. The model's coefficient of determination (R^2) was found to be 91.72% (0.9172), which was close to 1. The R^2 value (91.72%) implies that 91.72% of the variation in bioethanol production was due to the independent parameters. Overall, the model accounts for a significant portion of the variability in the response variable, with pH, Na_2HPO_4, and fermentation time playing significant roles individually and through their interactions.

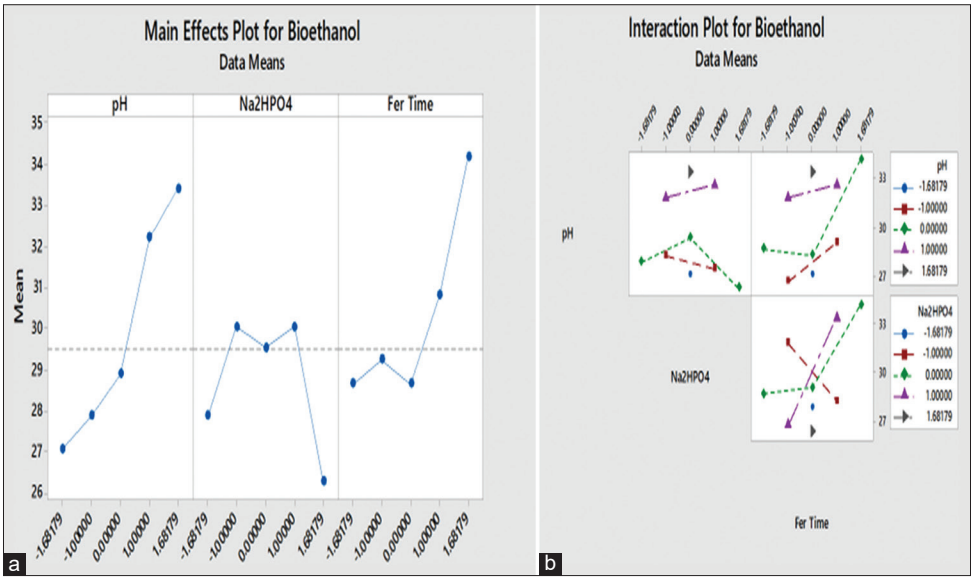

Figure 3: (a) Main effect plots (b) Interaction plots

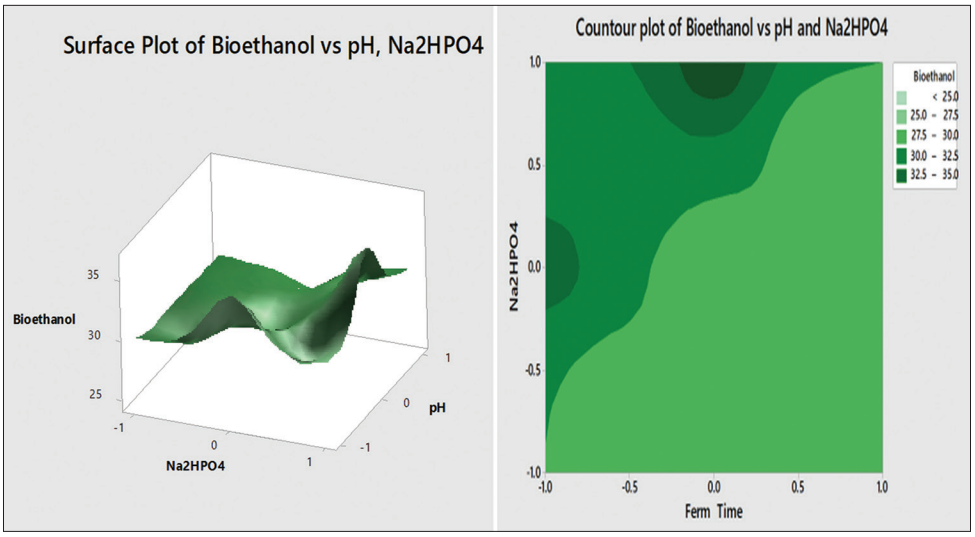

Figure 4: The 3D surface plots and 2D contour plot showing the relative effect of pH and Na 2 HPO 4 on bioethanol production.

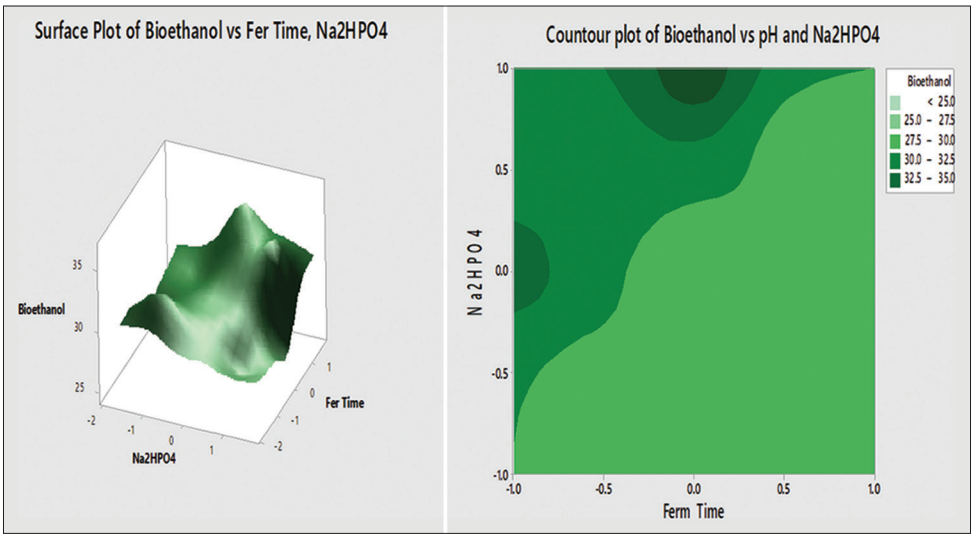

Figure 5: The 3D surface plots and 2D contour plot showing the relative effect of time and Na 2 HPO 4 on bioethanol production.

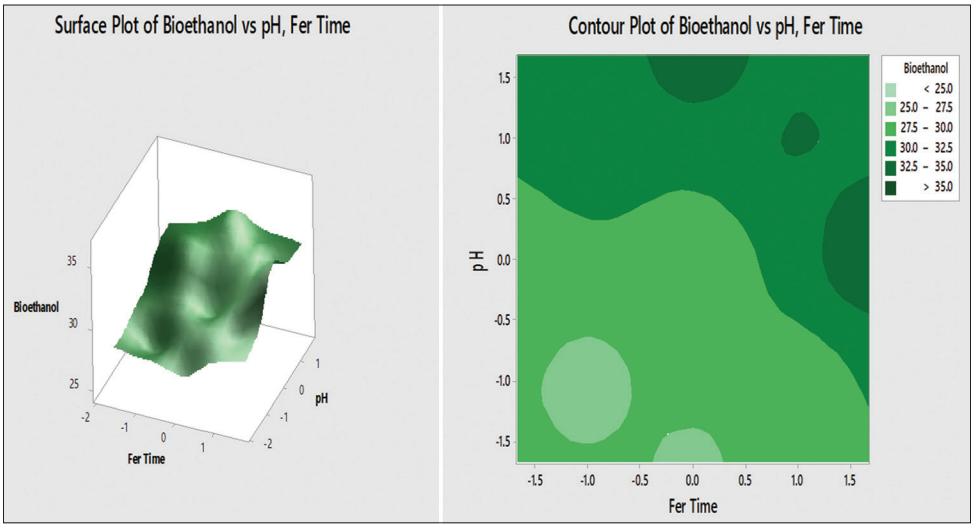

Figure 6: The 3D surface plots and 2D contour plot showing the relative effect of pH and Fer time on bioethanol production.

Table 2: Central composite design matrix for the production of bioethanol

Order	pH	Na_2HPO_4	F fermentation time	Expt results (g/L)	RSM predicted (g/L)
13	5.0	0.3	24	28.670	29.7392
8	5.5	0.4	60	36.540	35.4356
18	5.0	0.3	48	27.881	28.6540
17	5.0	0.3	48	27.881	28.6540
2	5.5	0.15	36	34.980	33.5534
11	5.0	0.047	48	27.880	27.6479
20	5.0	0.3	48	29.450	28.6540
19	5.0	0.3	48	28.670	28.6540
4	5.5	0.45	36	28.670	28.8249
3	4.5	0.45	36	24.720	23.1499
1	4.5	0.15	36	28.670	29.4535
12	5.0	0.55	48	26.300	26.9861
6	5.5	0.15	60	28.670	29.9191
7	4.5	0.45	60	30.240	31.3456
5	4.5	0.15	60	27.880	27.4041
10	5.8	0.30	48	33.400	33.9154
9	4.1	0.30	48	27.090	27.0285
15	5.0	0.30	48	28.670	28.6540
14	5.0	0.30	72	34.190	33.5748
16	5.0	0.30	48	29.450	28.6540

Table 3: Model summary table of RSM (CCD)

S	R-sq (%)	R-sq (adj) (%)	PRESS	R-sq (pred) (%)
1.18934	91.72	84.27	104.452	78.62

RSM: Response surface methodology, CCD: Central composite design, F: Fisher's function, DF: Degree of freedom, Adj. SS: Adjusted sum of squares, Adj MS: Adjusted mean squares. R^2=91.72%; Adjusted R^2=84.27%; Predicted R^2=78.62%. Any probability P<0.05 corresponds to significance

From the main effect plots (Figure 3), it was found that with an increase in pH the production of bioethanol was also increased and for Na_2HPO_4 until the middle value there was an increase in the bioethanol, production, the effect of fermentation time was a significant parameter as it was evidenced from different plots.

Figure 4; represents the relative effects of Na_2HPO_4 (0.05–0.55) and pH (4.1–5.8) on bioethanol production, while holding fermentation time constant at 48 h. The bioethanol production was high at a very high pH and mid value of Na_2HPO_4. The plots reveled that as an increase in pH and keeping Na_2HPO_4 value at mid-level, the bioethanol production was also increases.

Figure 5 above represents the relative effects of Na_2HPO_4 (0.05–0.55) and fermentation time (24–72h) on bioethanol production while holding a pH as 4. The bioethanol production was high at more fermentation time and mid value of Na_2HPO_4.

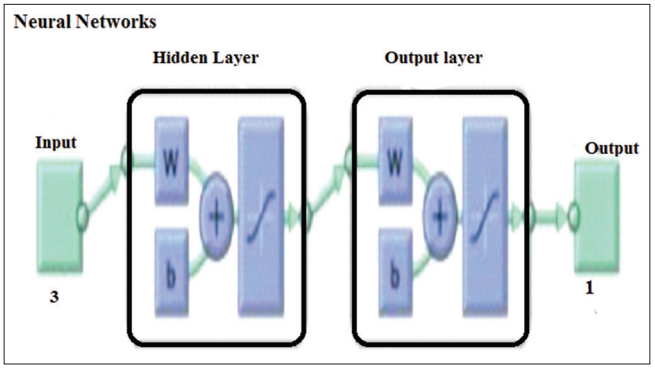

Figure 7: Artificial neural networks: Model (3–10–1)

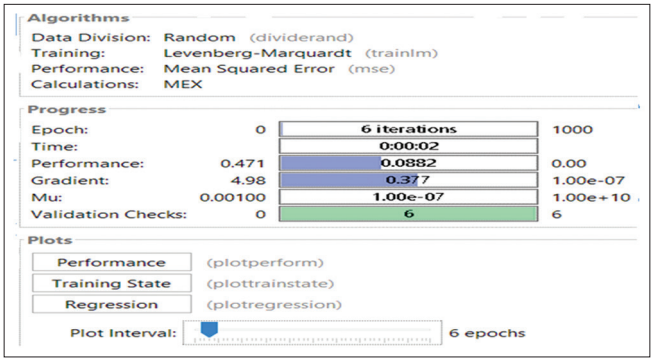

Figure 8: The type of algorithms used in prediction in artificial neural networks

Figure 6; illustrates the relative effects of fermentation time (24–72 h) and pH (4.1–5.8) on bioethanol production,

Table 4: ANN model for production of bioethanol

Order	pH	N_a2HPO_4	Fetim fermentation time	Expt. result (g/L)	ANN model predicted (g/L)
Training data					
13	5.0	0.3	24	28.670	28.66
8	5.5	0.4	60	36.540	36.43
18	5.0	0.3	48	27.881	28.66
17	5.0	0.3	48	27.881	28.66
2	5.5	0.15	36	34.980	34.98
11	5.0	0.047	48	27.880	27.88
20	5.0	0.3	48	29.450	28.66
19	5.0	0.3	48	28.670	28.66
4	5.5	0.45	36	28.670	28.70
3	4.5	0.45	36	24.720	24.73
1	4.5	0.15	36	28.670	28.67
12	5.0	0.55	48	26.300	26.29
6	5.5	0.15	60	28.670	28.67
7	4.5	0.45	60	30.240	30.24
5	4.5	0.15	60	27.880	27.86
Validation					
10	5.8	0.30	48	33.400	33.40
9	4.1	0.30	48	27.090	27.09
15	5.0	0.30	48	28.670	28.66
Testing					
14	5.0	0.30	72	34.190	34.19
16	5.0	0.30	48	29.450	28.66

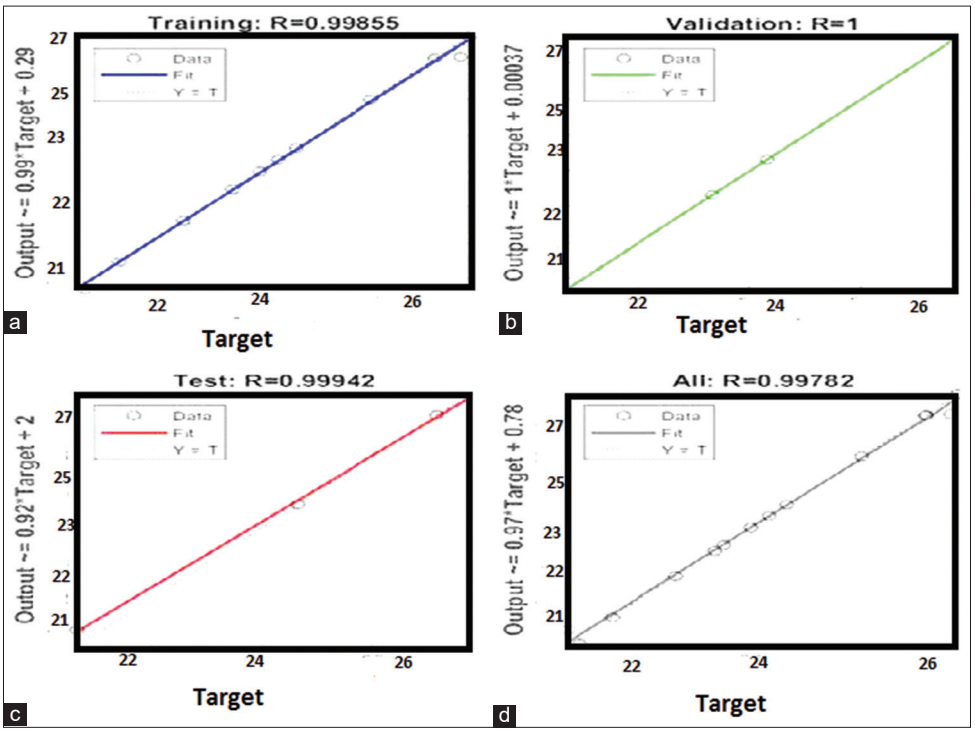

Figure 9: Regression plot observed versus predicted results for (a) training, (b) validation, (c) testing and (d) total, followed by their respective R^2 values

Table 5: Comparison of optimal condition prediction for RSM and ANN

Tool	pH	Na_2HPO_4	Fermentation time	Actual bioethanol yield (g/L)	Predicted bioethanol yield (g/L)
RSM (CCD)	5.5	0.4	60	36.54	35.43
ANN	5.5	0.4	60	36.54	36.45

RSM: Response surface methodology, ANN: Artificial neural networks, CCD: Central composite design

while keeping Na_2HPO_4 constant at 0.3. Bioethanol production was higher at increased values of fermentation time and pH. As both pH and fermentation time increases, bioethanol production also increased. To validate the model's adequacy, experiments were conducted in triplicate within the experimental range to verify the predicted optima. The experimental results concurred with the predicted values, confirming the model's adequacy. Optimal values for different parameters were obtained using the Minitab optimizer. The predicted bioethanol production at these optimal values was 35.43 g/L, and the experimental value was 36.54 g/L, aligning closely with the prediction.

ANN modeling and optimization

ANN modeling was implemented using the MATLAB® Neural Network Toolbox based on Haykin's methodology.[19] The input layer included normalized experimental variables: Fermentation time, pH, and Na_2HPO_4. The hidden layer consisted of 10 neurons, a number determined by testing up to 50 neurons and selecting the number that allowed the ANN to best learn and generalize the experiment (i.e., the smallest mean squared error [MSE] and highest R^2 values). The output layer had one neuron for estimating lipase production based on the input variables. Sigmoidal functions were used as activation functions in the hidden layer, and a linear function was used in the output layer. Additional parameters were kept at MATLAB's default settings. The training, validation, and test samples used in this study are detailed in a separate table. Out of a total of 20 samples, 15 were used for training (samples 13–16 were averaged values of the central point), 3 for validation, and 2 for testing. Although 20 samples are generally considered a small dataset for ANN training, the high quality of the predicted values (R^2 values >0.999) justified their adequacy due to the representativeness and precision of the data. All samples were averaged triplicates to minimize outlier influence. Data were generated using the CCD with two additional upper and lower levels, extending beyond the original experimental design's domain (Tables 2-4).

Optimal number of hidden neurons

Increasing the number of hidden neurons usually improves learning performance up to a certain point. Too few neurons can restrict the neural network's ability to model the process effectively, while too many can lead to overfitting, where the network learns noise present in the training data.[20]

The impact of varying the number of hidden neurons on model fit was evaluated, revealing that 10 hidden neurons provided the best balance. Using more neurons resulted in noticeable overfitting. Therefore, a 3–10–1 topology was selected as the optimal configuration for estimating bioethanol production (Figure 7).

Figure 8 has information of the type of data division, which is random and the training equation is Levenberg-Marquardt and here the performance type is MSE.[21] The Levenberg-Marquardt training algorithm was a precise optimization method commonly used in neural network training. In MATLAB, the "nntool" function referred to a graphical user interface for training neural networks with various algorithms, including Levenberg-Marquardt. This algorithm was particularly popular for solving non-linear regression problems as it combined the advantages of the Gauss-Newton method and the gradient descent method (Figure 9). It efficiently handled highly non-linear mappings and often converged faster than traditional gradient-based optimization algorithms.[22]

Comparison of RSM and ANN predicted values

Comparing the predicted and actual bioethanol output values from RSM and ANN, both models demonstrated strong performance based on R^2 and AAD values, providing consistent responses. However, the ANN approach outperformed RSM in terms of both data fitting and estimation capabilities.[16,23,24]

CONCLUSION

This research study confirms that agricultural waste Areca nut leaves were used to produce bioethanol by the separate hydrolysis and fermentation methods. During pre-treatment with acid at high-temperature plant cell walls will be disrupted and in the enzyme hydrolysis process using cellulose enzyme we convert cellulose into glucose units and in the yeast fermentation process we convert sugar into bioethanol, we purify the bioethanol using rotary evaporator based on the boiling point of bioethanol. Here, the fermentation process is optimized by ANN and RSM. ANNs as compared to RSM were successfully applied to the optimization and prediction of bioethanol production. The high regression coefficients R^2 and the low root mean square error of the ANN model revealed that it was well fitted to the experimental design. Hence, the results of the significance levels were found to be pH, fermentation time and Na_2HPO_4 were the most significant factors affecting the bioethanol concentrations from the fermentation process. The optimal conditions were

pH 5.5, 60 h of fermentation, and 0.45 g of Na_2HPO_4, under these optimal conditions we get 36.54 g/L bioethanol yield.

ACKNOWLEDGMENT

The authors are thankful to KLE Technological University, Hubbali, Karnataka, India and ISNC, Jeddah, KSA, and Administration for their constant support.

REFERENCES

1. Antunes FA, Chandel AK, Terán-Hilares R, Milessi TS, Travalia BM, Ferrari FA, *et al*. Biofuel production from sugarcane in Brazil. In: Sugarcane Biofuels: Status, Potential, and Prospects of the Sweet Crop to Fuel the World. Germany: Springer; 2019. p. 99-121.
2. Hossain N, Razali AN, Mahlia TM, Chowdhury T, Chowdhury H, Ong HC, *et al*. Experimental investigation, techno-economic analysis and environmental impact of bioethanol production from banana stem. Energies 2019;12:3947.
3. De Souza WR, De Gouvea PF, Savoldi M, Malavazi I, De Souza Bernardes LA, Goldman MH, *et al*. Transcriptome analysis of *Aspergillus niger* grown on sugarcane bagasse. Biotechnol Biofuels 2011;4:1-7.
4. Gamage J, Lam H, Zhang Z. Bioethanol production from lignocellulosic biomass, a review. J Biobased Mater Bioenergy 2010;4:3-11.
5. Hossain N, Mahlia TM, Miskat MI, Chowdhury T, Barua P, Chowdhury H, *et al*. Bioethanol production from forest residues and life cycle cost analysis of bioethanol-gasoline blend on transportation sector. J Environ Chem Eng 2021;9:105542.
6. Hossain N, Zaini JH, Mahlia TM. A review of bioethanol production from plant-based waste biomass by yeast fermentation. Int J Technol 2017;8:5-18.
7. Betiku E, Taiwo AE. Modeling and optimization of bioethanol production from breadfruit starch hydrolyzate vis-à-vis response surface methodology and artificial neural network. Renew Energy 2015;74:87-94.
8. Sebayang IS, Suroso A, Laoli AG. Optimization rainfall-runoff modeling for Ciujung river using back propagation method. SINERGI 2018;22:193-204.
9. Chouaibi M, Daoued KB, Riguane K, Rouissi T, Ferrari G. Production of bioethanol from pumpkin peel wastes: Comparison between response surface methodology (RSM) and artificial neural networks (ANN). Ind Crops Prod 2020;155:112822.
10. Kamairudin N, Abd Gani SS, Masoumi HR, Basri M, Hashim P, Mokhtar NM, *et al*. Modeling of a natural lipstick formulation using an artificial neural network. RSC Adv 2015;84:68632-8.
11. Zou J, Han Y, So SS. Overview of artificial neural networks. Methods Mol Biol 2009;458:14-22.
12. Kiani MK, Ghobadian B, Tavakoli T, Nikbakht AM, Najafi G. Application of artificial neural networks for the prediction of performance and exhaust emissions in SI engine using ethanol-gasoline blends. Energy 2010;35:65-9.
13. Najafi G, Ghobadian B, Tavakoli T, Buttsworth DR, Yusaf TF, Faizollahnejad MJ. Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network. Appl Energy 2009;86:630-9.
14. Ayodele TR, Ogunjuyigbe AS, Monyei CG. On the global solar radiation prediction methods. J Renew Sustain Energy 2016;8:023702.
15. Tasdemir S, Saritas I, Ciniviz M, Allahverdi N. Artificial neural network and fuzzy expert system comparison for prediction of performance and emission parameters on a gasoline engine. Expert Syst Appl 2011;38:13912-23.
16. Raina N, Slathia PS, Sharma P. Response surface methodology (RSM) for optimization of thermochemical pretreatment method and enzymatic hydrolysis of deodar sawdust (DS) for bioethanol production using separate hydrolysis and co-fermentation (SHCF). Biomass Convers Biorefin 2020;28:1-21.
17. Sherpa KC, Ghangrekar MM, Banerjee R. Optimization of saccharification of enzymatically pretreated sugarcane tops by response surface methodology for ethanol production. Biofuels 2019;10:73-80.
18. Bušić A, Marđetko N, Kundas S, Morzak G, Belskaya H, Ivančić Šantek M, *et al*. Proizvod njabi oetanola izobnovlj ivihsirovi natenjego voodvajanjeipročišćavanje: pregled. Food Technol Biotechnol 2018;56:289-311.
19. Díaz-Moreno P, Carrasco JJ, Soria-Olivas E, Martínez-Martínez JM, Escandell-Montero P, Gómez-SanchisJ. Educational software based on matlab GUIs for neural networks courses. In: Handbook of Research on Computational Simulation and Modeling in Engineering. United States: IGI Global; 2016. p. 333-58.
20. Linko S, Luopa J, Zhu YH. Neural networks as 'software sensors' in enzyme production. J Biotechnol 1997;52:257-66.
21. Hombalimath VS, Gurumurthy DM. Response surface methodology (RSM) and artificial neural network (ANN) integrated optimization for lipase production by Bacillus holotolerans. Syst Microbiol Biomanufact 2024;3:1140-9.
22. Moré JJ. The Levenberg-Marquardt algorithm: Implementation and theory. In: Numerical Analysis: Proceedings of the Biennial Conference held at Dundee, June 28-July 1, 1977. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006. p. 105-16.
23. Pereira LM, Milan TM, Tapia-Blácido DR. Using response surfaceMethodology (RSM) to optimize 2G bioethanol production: A review. Biomass Bioenergy 2021;151:106166
24. Jang R, Chous YS. Methodology (RSM) and artificial neural networks (ANN). Ind Crops Prod 20202;155:112822.

Source of Support: Nil. **Conflicts of Interest:** None declared.