Predictive Modeling of Agricultural Biomass Paper Production Using Machine Learning Approaches

Dharunesh Balasubramaniam Gopalakrishnan., Divya Nithiya., Priyanka Murali., Prathiba Rex[*]

Department of Biotechnology, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai 62, Tamil Nadu, India

[*]Corresponding Author

DOI: https://dx.doi.org/10.51584/IJRIAS.2026.11050190

Received: 17 May 2026; Accepted: 22 May 2026; Published: 13 June 2026

ABSTRACT

The transformation of biomass into high-value materials presents a promising solution for sustainable development. Accurate prediction of material properties based on biomass composition and processing parameters is critical for optimizing production efficiency and product quality. This study explores the application of multiple machine learning algorithms— Linear Regression, Decision Tree, Random Forest, and Support Vector Regression (SVR)—to predict material characteristics derived from biomass inputs such as cellulose, lignin, hemicellulose content, pulping time, and energy consumption. A dataset based on experimental values was used to train and evaluate the models. Among the tested approaches, the decision tree model showed the highest prediction performance (R^2 = 1.000); however, the limited dataset size may have contributed to possible overfitting followed by Linear Regression (R^2= 0.987) and Random Forest (R^2 = 0.976), while SVR showed limited performance (R^2 = 0.081) due to the small dataset size. The results highlight the effectiveness of tree-based and linear models in accurately modeling the complex interactions between biomass composition and processing parameters. This research underscores the potential of machine learning techniques in advancing biomass valorization strategies and optimizing sustainable material production. The study demonstrates the feasibility of machine learning-assisted prediction for biomass-based paper production, while highlighting the need for larger datasets and industrial-scale validation

Keywords: Sustainable Paper Production, Agricultural Biomass Residues, Sugarcane Bagasse, Corn Stalks, Pulping Methods, Cellulose Extraction

INTRODUCTION

Paper is indispensable in daily life, with applications ranging from writing and printing to packaging and photography. Globally, approximately 300 million tonnes of paper are produced annually, predominantly from fibrous wood. This dependency on wood exacerbates environmental issues such as deforestation and climate change. Agricultural residues, a largely untapped resource, present a promising alternative. Currently, only 8% of global paper production utilizes agricultural wastes, despite their abundance and cost-effectiveness. A comprehensive overview of the potential of agricultural residues as alternative raw materials in the pulp and paper industry was studied [1]. The process of manufacturing paper using agricultural residues as the primary raw material and the article covers the technical aspects of converting various residues like bagasse, straw, and husks into paper products, including the chemical and mechanical treatments involved [2]. Fahmy et al. [3] explores the pre-extraction of reducing sugars from wheat straw before its use in pulp and paper production. The research demonstrates that this pre-treatment enhances the pulping process by improving fiber quality and reducing the need for harsh chemicals. A comprehensive background on the utilization of agricultural residues for paper and board production was investigated by Ciambelloti et al. [4] and discusses various types of agricultural wastes and their potential for use in the paper industry, as well as future prospects and challenges. This study investigates the use of agricultural biomass residues for sustainable paper production, focusing on optimizing the extraction of cellulosic components for nanocellulose and paper manufacturing. To manufacture paper from agricultural biomass residues, reducing environmental impact and providing an alternative to traditional wood-based paper production. Effectively collect and repurpose agricultural residues (sugarcane bagasse, corn stalks, paddy straws, coconut fiber, and husks) as raw materials for paper production. Develop a pulping method that ensures the final product meets industry standards for strength, texture, and durability. Minimize the environmental footprint by reducing harmful chemicals and energy use while promoting renewable resources.

The pulp and paper industry heavily depends on fossil fuels, with tissue manufacturing requires high thermal and electrical energy. Researchers showed that integration of bioenergy systems can reduce CO_2 emissions by 12 %–80 %. However, full biomass replacement faces logistical and economic challenges. Partial substitution offers a more feasible path for decarbonization [5]. Adeleke et al. [6] prepared a biomass derived carbon alloy catalyst nano-blood charcoal (NBC), from blood meal and cellulose nanofibers, enabled the development of a high-perfor-mance water-activated Mg–air paper battery. The battery achieved an open-circuit voltage of 1.57 V, a maximum current density of 161 mA/cm², and a power density of 55.7 mW/cm².

NBC also showed good oxygen evolution reaction (OER) performance, supporting rechargeable functionality.

The objective of this study is to develop machine learning models for predicting paper quality from agricultural biomass residues based on biomass composition and processing parameters. Cellulose, lignin, hemicellulose content, pulping time, and energy consumption were used as input variables. Four machine learning algorithms, namely Linear Regression, Decision Tree, Random Forest, and Support Vector Regression, were comparatively evaluated using R^2 and MSE metrics to identify the most suitable predictive model for sustainable biomass-based paper production.

1. Experimental

Collection of agricultural waste, preprocessing, mechanical pulping, screening and purification, beating and refining of pulp and paper-making process. Agricultural residues such as sugarcane bagasse, corn husk and coconut sheath were collected and soaked in water for 24-48 hours to facilitate softening shown in Fig 1. The softened residues were ground to reduce particle size and prepared them for pulping. The ground materials underwent chemical pulping to extract cellulose fibers. Screening and purification processes were carried out to remove impurities, followed by beating and refining to enhance pulp quality. The refined pulp was transformed into paper sheets through conventional paper-making techniques.

Fig 1. Agricultural waste (a) sugarcane bagasse (b) corn husk and (c) coconut sheath

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue V May 2026

Energy Consumption:

$E = p \times t$ (1)

Where E is the energy consumption (kWh), p is the power rating of equipment (kW) and t is the operating time (hr).

Paper Quality Prediction (Regression Model):

$Q = \beta 0 + \beta 1C + \beta 2L + \beta 3H + \beta 4T + \beta 5E + \epsilon$(2)

Where Q is the paper quality score, C is the cellulose content (%), L is the lignin content (%), H is the hemicellulose content (%), T is the chemical pulping time (hr), E is the energy consumption (kWh), β are regression coefficients and ϵ is the error term.

The dataset consisting of 100 experimental samples was divided into training and testing sets using an 80:20 ratio. The machine learning models were implemented in Python using Scikit-learn libraries within the Google Colab environment. Linear Regression was used as a baseline model, while Decision Tree and Random Forest algorithms were applied to capture nonlinear relationships. Support Vector Regression with radial basis function (RBF) kernel was also evaluated. Model performance was assessed using Mean Squared Error (MSE) and coefficient of determination (R^2). Due to the limited dataset size, the obtained results should be considered preliminary and intended for comparative analysis rather than industrial-scale prediction.

In this study, four machine learning (ML) models were used to study the effectiveness of paper quality. These include a Linear Regression (LR) model, a support vector regression (SVR) method, and two tree-based algorithms—decision tree (DT) and random forest (RF). These models were selected based on their established predictive capabilities in biomass-based modeling and materials optimization applications [7]. The models used five independent variables derived from biomass and process conditions, namely cellulose content (%), lignin content (%), hemicellulose content (%), pulping time (hours), energy consumption (kWh) and the target variable is the predicted material property value (quality score). Google Colab environment was used to perform the algorithms with python coding.

Model Implementation

- Linear Regression (LR): A baseline model that estimates the target as a weighted linear combination of the input features. It was used to provide interpretability and understand feature contributions via regression coefficients.

- Support Vector Regression (SVR): Implemented using the radial basis function (RBF) kernel to handle potential nonlinearities. It attempts to fit the best possible function within a margin of tolerance, making it robust to outliers.

- Decision Tree (DT): A non-parametric model that splits the data based on feature thresholds to minimize prediction error. It is easy to interpret and useful for identifying decision rules.

- Random Forest (RF): An ensemble learning method consisting of multiple decision trees. It improves generalization by averaging predictions, reducing overfitting, and handling feature interactions effectively.

All models were trained using a supervised learning approach on a dataset of 100 samples. Predictions were evaluated using Mean Squared Error (MSE) and R-squared (R^2) as performance metrics [8].

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue V May 2026

RESULTS AND DISCUSSION

Characterization of biomass

FTIR spectra for three different biomass coconut sheath (CS), corn husk (CH), and sugarcane bagasse (SB) is shown in Fig 2. Broad band near 3300–3500 cm^{-1} indicates O–H stretching vibrations, commonly found in hydroxyl groups (cellulose or polysaccharides). Peak at 2900 cm^{-1} indicates C–H stretching (aliphatic compounds). Peak near 1050–1150 cm^{-1} shows the C–O stretching vibrations in alcohols or ethers. Strong band near 1000 cm^{-1} is due to C–N or other ring structures, indicates presence of potential proteins or amines. Peaks around 1600–1700 cm^{-1} is C=C (alkene) or amide C=O groups, which has some proteinaceous or aromatic structures. Sugarcane bagasse has rich functional groups other than CH and CS.

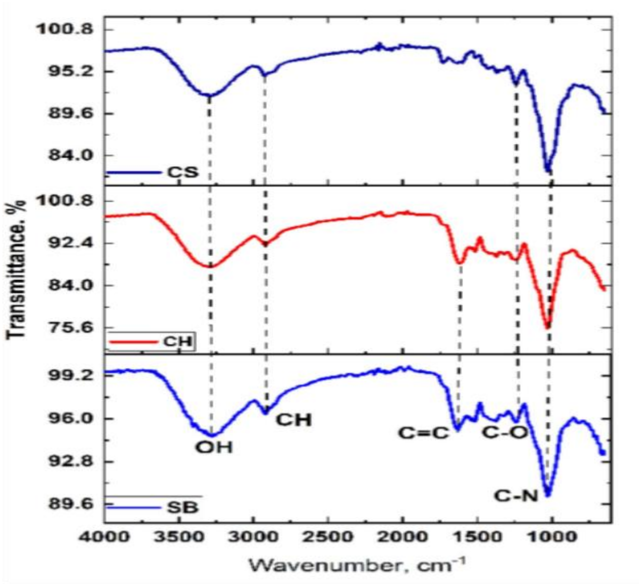

Fig 2. FTIR of sugarcane bagasse (SB), corn husk (CH) and coconut sheath (CS)

Formation of Biomass Paper

The preliminary experiments demonstrated that agricultural residues are effective raw materials for paper production. Sugarcane bagasse and corn stalks exhibited high cellulose content, which facilitated the pulping process [9]. The chemical pulping method resulted in pulp with adequate strength and fiber length, meeting the baseline requirements for paper production. Mechanical and chemical processes were optimized to minimize waste and improve efficiency. Fig 1 shows the sugarcane bagasse and corn stalk soaked for 24 hr. Screening and purification ensured that impurities were effectively removed, resulting in cleaner pulp. The final paper sheets displayed promising physical properties, including:

- Tensile Strength - Comparable to traditional wood-based paper.
- Texture and Appearance - Smooth surface with uniform texture.
- Durability - Adequate for writing and printing applications [10].

The findings confirm that agricultural residues such as sugarcane bagasse and corn stalks are viable substitutes for wood in paper production. Fig 3 shows the grinded materials of (a) sugarcane bagasse (b) corn husk (c) coconut sheath. The high cellulose content and low lignin levels in these materials contributed to their suitability. Fig 4 shows the wood paper produced from pulping method. By leveraging agricultural waste, this approach addresses two critical challenges: reducing reliance on wood and managing agricultural residues sustainably. However, there are areas for improvement. For instance, the chemical pulping process, while effective, relies on chemicals that could pose environmental risks. Future efforts should focus on integrating eco-friendly alternatives such as enzymatic pulping. Additionally, the mechanical pulping process could be further refined to

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue V May 2026

enhance fiber quality and reduce energy consumption. Scaling up production will require addressing logistical challenges in collecting and transporting agricultural residues [11]. A decentralized collection and preprocessing system could mitigate these issues and enhance the economic feasibility of the process. Finally, conducting life cycle assessments will provide a clearer picture of the environmental benefits and potential trade-offs associated with this approach. Table 1 shows the quality of paper and energy consumption from different biomass.

Fig 3. Ground materials of (a) sugarcane bagasse (b) corn husk (c) coconut sheath

Fig 4. Formation of biomass paper (a) Wooden Mold for biomass paper and (b) Biomass paper

Table 1 quality of paper and energy consumption from different biomass

Cellulose (wt.%)	Lignin (wt.%)	Hemicellulose (wt.%)	Chemical pulping Time (hr)	Energy consumption (kWh)	Paper quality (%)
47.49	5.31	26.42	2.21	60.31	27.90
59.01	11.36	20.84	4.13	140.26	33.61
54.64	8.14	21.62	4.16	100.53	30.68
51.97	10.09	28.99	4.55	132.65	27.15
43.12	14.08	26.06	4.90	82.00	20.26

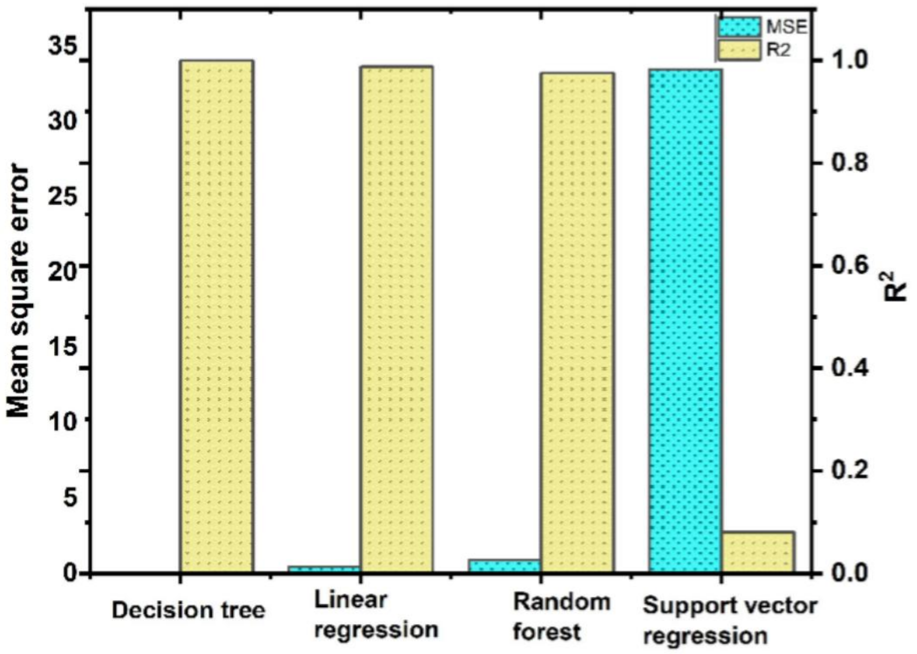

Fig 5. Mean square error and R^2 for different algorithms

CONCLUSIONS

This study successfully demonstrated the application of machine learning algorithms to predict material characteristics based on biomass composition and processing parameters. By comparing four different models—Linear Regression, Decision Tree, Random Forest, and Support Vector Regression—valuable insights into model performance were obtained. Tree-based models, particularly the Decision Tree and Random Forest, exhibited superior prediction accuracy, while Support Vector Regression showed limited applicability for the current dataset size. The findings affirm that machine learning approaches can effectively model complex, multivariable relationships in biomass-based material development. These predictive models offer a powerful tool for optimizing biomass valorization processes, reducing experimental costs, and supporting the advancement of sustainable material production. Future work will focus on expanding the dataset, incorporating feature selection techniques, and validating the models under industrial-scale conditions to further enhance predictive performance and real-world applicability.

REFERENCES

1. L. A. Worku, A. Bachheti, R. K. Bachheti, C. E. Rodrigues Reis, and A. K. Chandel, "Agricultural Residues as Raw Materials for Pulp and Paper Production: Overview and Applications on Membrane Fabrication," *Membranes (Basel).*, vol. 13, no. 2, pp. 1–17, 2023, doi: 10.3390/membranes13020228.
2. S. Malik, V. Rana, G. Joshi, P. K. Gupta, and A. Sharma, "Valorization of Wheat Straw for the Paper Industry: Pre-extraction of Reducing Sugars and Its Effect on Pulping and Papermaking Properties," *ACS Omega*, vol. 5, no. 47, pp. 30704–30715, 2020, doi: 10.1021/acsomega.0c04883.
3. Y. Fahmy, T. Y. A. Fahmy, F. Mobarak, M. El-Sakhawy, and M. H. Fadl, "Agricultural Residues (Wastes) for Manufacture of Paper, Board, and Miscellaneous Products: Background Overview and Future Prospects," *Int. J. ChemTech Res.*, vol. 10, no. 2, pp. 424–448, 2017, doi: 10.5281/zenodo.546735.
4. A. Ciambellotti, A. Baldinelli, S. Frigo, S. Pecchia, and U. Desideri, "Energy transition in tissue paper industry through solid biomass combustion and gasification: Techno- economic potential and limitations," *J. Clean. Prod.*, vol. 495, p. 145014, 2025, doi: https://doi.org/10.1016/j.jclepro.2025.145014.
5. K. Ishibashi and H. Yabu, "Sustainable water-activated metal–air paper batteries based on waste biomass-

based electrocatalysts††Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4su00607k," *RSC Sustain.*, vol. 2, no. 12, pp. 4046–4051, 2024, doi: https://doi.org/10.1039/d4su00607k.

6. A. A. Adeleke *et al.*, "Comparative studies of machine learning models for predicting higher heating values of biomass," *Digit. Chem. Eng.*, vol. 12, no. May, p. 100159, 2024, doi: 10.1016/j.dche.2024.100159.
7. V. R. Srenyvasan, S. Muthuvel, M. K. Rahiman, S. S. Khan, and P. Rex, "Journal of Water Process Engineering Zinc manganese oxide spinel phase nanomaterial layered on solar stills applications : An analytical study for improved performance compared with machine learning methods," *J. Water Process Eng.*, vol. 71, no. November 2024, p. 107363, 2025, doi: 10.1016/j.jwpe.2025.107363.
8. P. Rex and K. Rahiman, "Synergistic valorization of wheat husk-derived HZSM-5 catalyst in pyrolysis of polystyrene and polypropylene: sustainable waste-to-energy conversion enhanced by machine learning models," *J. Mater. Cycles Waste Manag.*, no. 0123456789, 2024, doi: 10.1007/s10163-024-02048-9.
9. H. M. Kadlimatti, B. Raj Mohan, and M. B. Saidutta, "Microwave-assisted pyrolysis of food waste: optimization of fixed carbon content using response surface methodology," *Biofuels*, vol. 12, no. 9, pp. 1051–1058, 2021, doi: 10.1080/17597269.2019.1573609.
10. P. Rex, N. Meenakshisundaram, and P. Barmavatu, "Sustainable valorisation of kitchen waste through greenhouse solar drying and microwave pyrolysis– technology readiness level for the production of biochar," *J. Environ. Heal. Sci. Eng.*, p. 40201, 2024, doi: 10.1007/s40201-024-00909-x.
11. P. Rex, K. Rahiman, M. Ismail, N. Meenakshisundaram, P. Barmavatu, and A. V. S. L. S. Bharadwaj, "Agricultural Biomass Waste to Biochar : A Review on Biochar Applications Using Machine Learning Approach and Circular Economy," 2023.