Development Of Biochar Yield Prediction Model From Food Waste Pyrolysis

2024 19th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP) | 979-8-3315-0991-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/iSAI-NLP64410.2024.10799410

Yogan Jaya Kumar
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka
Melaka, Malaysia
yogan@utem.edu.my

Surend Esvaran
Faculty of Information and Communication Technology
Universiti Teknikal Malaysia Melaka
Melaka, Malaysia
esurend@gmail.com

Siti Fatimah Zaharah M Fuzi
Faculty of Sciences & Technology
Universiti Tun Hussein Onn Malaysia
Johor, Malaysia
fatimahz@uthm.edu.my

Nasratun Masngut
Faculty of Chemical and Process Engineering Technology
Universiti Malaysia Pahang Al-Sultan Abdullah
Pahang, Malaysia
nasratun@umpsa.edu.my

Fathin Ayuni Azizan
Faculty of Mechanical Engineering & Technology
Universiti Malaysia Perlis
Perlis, Malaysia
fathin@unimap.edu.my

***Abstract*—The accumulation of waste in natural surroundings has dramatically increased, necessitating new, sustainable solutions for food waste management. Food waste, a significant environmental challenge, can be transformed into biochar. As a carbon-rich substance formed from biomass pyrolysis, biochar has garnered interest for its potential as a sustainable soil amendment, carbon sequestration technique, and renewable energy source. However, biochar yield from food waste pyrolysis is influenced by parameters such as pyrolysis condition and feedstock composition. Accurate biochar yield prediction is crucial for effective food waste management, optimizing resource utilization, and minimizing environmental impact. Artificial intelligence (AI) algorithms have shown potential in modeling complex systems and predicting outcomes. This study aims to investigate the performance of AI-driven models for predicting biochar yield from food waste pyrolysis. Models such as Linear Regression, Random Forest, K-Nearest Neighbors, and Convolutional Neural Networks (CNNs) are evaluated. Data preprocessing, specifically feature scaling and logarithmic transformation, were found to play a pivotal role in enhancing the model performance. These advancements will contribute to more sustainable biochar production.**

***Keywords*—biochar, food waste, pyrolysis, artificial intelligence, prediction model**

I. INTRODUCTION

Food waste management is a critical global environmental and economic issue. According to the Food Waste Index Report 2024 by the United Nations Environment Programme (UNEP), in 2022, the world wasted 1.05 billion tonnes of food. Food loss and waste generates 8 to10 percent of global greenhouse gas emissions. These emissions arise from every stage of the food supply chain. Food waste also comprises biodegradable waste coming from sources such as food processing and catering industries [1].

The most common approach for disposing food waste is through landfilling. Nevertheless, this conventional method has its sustainability challenges, leading to the release of hazardous gases, contamination of water and soil, and extensive land occupation [2], [3]. The persistent degradation of the environment and the gradual depletion of fossil fuel reserves have intensified research on food waste conversion into sustainable and renewable resources [4].

Biomass, a renewable energy source, can be generated from food waste, through thermal decomposition of materials at elevated temperatures. This thermal decomposition process can transform food waste into valuable biochar, a carbon-rich substance formed from biomass pyrolysis.

Pyrolysis , a thermal decomposition method, has been used as a sustainable method for treating organic waste, yielding a range of valuable products such as heat, electricity, biogas, biochar, and bio-oil. The process involves the degradation of organic molecules in an oxygen-deprived environment, achieved by breaking chemical bonds within a moderate temperature range of 300–800 °C. It is worth noting that operational parameters, including temperature, residence time, heating rate, as well as the type and flow rate of inert gas, have a significant influence on the dynamics of the pyrolysis process and the resultant distribution and properties of its products [5].

Studies have shown that food waste pyrolysis is a promising method for biochar production. Biochar constitutes residual solid material obtained in a charred form, encompassing carbon content, inorganic materials, as well as metals and transition elements. This byproduct, arising from the process of food waste pyrolysis, have multiple uses; as a catalyst, adsorbent, and soil conditioner. Additionally, biochar serves as a significant agent for carbon sequestration, potential to mitigate the adverse effects of climate change and its associated consequences [6].

The performance of biochar in diverse applications is linked to its physical and chemical attributes, encompassing factors such as ash content, moisture levels, nutrient composition, and metal content [7]. Thus, estimating how much biochar can be produced from a given amount of food waste is needed to expand its applicability and economic value. Furthermore, customization and optimization of pyrolysis parameters are essential to enhance the properties of biochar, making it better suited for a targeted application.

However, predicting biochar yield is complex due to the numerous variables involved in the pyrolysis process. Therefore, this study develops AI-based predictive models to accurately estimate biochar yield from food waste pyrolysis, thereby promoting sustainable waste management.

II. Related Works

Past studies on renewable energy have been explored using innovative technologies and methodologies to enhance energy efficiency and sustainability. Renewable energy can be obtained from diverse sources, such as solar power, wind energy, biomass energy, geothermal energy and even radio-frequency energy harvesting [8]-[10], [19], [20]. These studies highlight the diverse applications and significant potential of renewable energy technologies in various fields, paving the way for more sustainable solutions.

Previous research works have effectively harnessed data-driven techniques utilizing machine learning algorithms for modeling various processes in generating renewable energy from biomass [11]. For instance, Hai et al. [12] applied an artificial neural network to model the pyrolysis of groundnut shell. Their evaluation of the Artificial Neural Network (ANN) model yielded a mean absolute error (MAE) of 0.786, highlighting its capability in predicting biochar production from groundnut. In a similar work, Tee et al. [13] employed ANN to establish a model linking different operating conditions to biochar yield. The ANN, augmented with the Levenberg-Marquardt algorithm, demonstrated superior robustness in predicting biochar yield.

In other studies, Pathy et al. [14] ventured into the domain of pyrolysis data associated with algal biomass. They developed a biochar prediction model employing an extreme gradient boosting (EGB) algorithm. The resultant model exhibited effectiveness in estimating biochar yield from the conversion of algal biomass.

Khan et al. [15] developed an integrated framework of ANN and metaheuristic algorithms for modeling the prediction of biochar from various biomass feedstocks, such as herbaceous plants, sewage sludge and animal manures. A comprehensive analysis considering the combined influence of proximate and ultimate analyses of agricultural biomasses, with the critical factor of pyrolysis conditions is still lacking and poorly understood [16].

Understanding the performance of different predictive models affecting biochar yield can lead to the optimization of pyrolysis processes, which can be beneficial for both waste reduction and biochar production.

III. Predictive Models

To address the problem of predicting biochar yield from food waste pyrolysis, various AI algorithms can be leveraged. Four AI-driven algorithms were chosen based on their diverse strengths and potential for regression tasks. Each model was trained using additional procedure such as feature scaling and logarithmic transformation to enhance performance.

A. Linear Regression Model

Linear regression, a statistical method that models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. Linear regression is chosen for its simplicity and interpretability in modeling relationships between biochar yield and input variables such as feedstock properties and pyrolysis conditions. By providing a straightforward and interpretable model, it allows for initial insights into the relationships between variables, useful for simple predictions and identifying key factors.

B. Random Forest Regressor

Conversely, the random forest regressor, is an ensemble learning method. This ensemble learning approach aggregates predictions from multiple decision trees, addressing overfitting concerns and enhancing predictive accuracy. For predicting biochar yield from food waste pyrolysis, random forest regressor can handle the complex, non-linear relationships between the input variables and the output. Its ability to manage a large number of input variables and interactions makes it suitable for the diverse and intricate nature of pyrolysis data.

C. K-Nearest Neighbors (KNN)

K-nearest neighbors (KNN), a non-parametric, instance-based learning is chosen for its capability to capture non-linear relationships between input variables such as feedstock properties and pyrolysis conditions, and biochar yield. Operating on the principle of similarity, KNN predicts the output for a new data point based on the most similar training examples. This adaptability allows KNN to effectively model complex and irregular data patterns in predicting biochar yield from food waste pyrolysis, scenarios where linear models may struggle. In the regression context, KNN predicts a value by averaging the values of its k nearest neighbors.

D. Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) are a type of deep learning model designed to process data with grid-like topology, such as images. Insights from Lee et al. [17] highlight their efficacy in processing multidimensional data such as that involved in biochar production. They consist of convolutional layers that apply filters to input data to capture spatial hierarchies and patterns, followed by pooling layers that reduce dimensionality and fully connected layers that perform the final regression. For predicting biochar yield from food waste pyrolysis, 1-dimensional CNNs can be adapted to handle structured data capturing multiple aspects of the pyrolysis process.

IV. Experimental Setup

The experimental setup of this study encompasses three critical components: data collection, model selection and training, and performance measurement.

A. Data Collection

Datasets from past studies focused on biochar production and pyrolysis were collected [18]. These data sources encompass food and agricultural waste data from pyrolysis experiments. These experiments provide crucial insights into variables such as Fixed carbon, Volatile matter, Ash content, Carbon (C), Hydrogen (H), Oxygen (O), Nitrogen (N), Residence time (min), Temperature (°C), Heating rate (°C/min), and Biochar yield. The input features comprise the first 10 variables, while the output is the Biochar yield. The total number of data is 226 samples.

B. Model Selection and Training

Four AI models, Linear Regression, Random Forest Regressor, K-Nearest Neighbors (KNN), and Convolutional Neural Networks (CNN) were chosen. Following model selection, the data was split into training (80%) and testing sets (20%). Each model was trained using additional procedure, incorporating data preprocessing steps such as feature scaling and logarithmic transformation to enhance performance. Fig. 1 shows the overall workflow for predicting biochar yield.

For Linear Regression, default parameters were used. The Random Forest Regressor was set with 100 trees. The KNN model utilized 2 neighbors. The CNN model included two Conv1D layers with 32 and 64 filters respectively, a kernel size of 3, and ReLU activation functions, followed by MaxPooling1D layers, a dense layer with 64 units, and an output layer with a linear activation function.

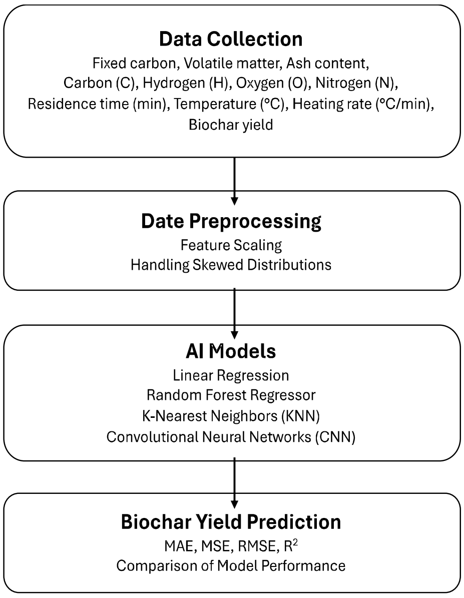

Fig. 1. Workflow for predicting biochar yield using AI models.

C. Performance Measurement

In developing a biochar yield prediction model using machine learning algorithms, it is essential to select appropriate performance metrics to evaluate and analyse the models' effectiveness. The performance of the models can be measured using several metrics, each providing insights into different aspects of the model's predictive capabilities and reliability. Here, we describe the performance measurements used in this study.

1) Mean Absolute Error (MAE): Mean Absolute Error measures the average magnitude of errors between predicted and actual values without considering direction. It is the mean of absolute differences, providing a straightforward measure of prediction accuracy. Lower MAE values indicate better model performance, making it useful for evaluating regression models.

2) Mean Squared Error (MSE): Mean Squared Error is the average of the squares of the differences between the predicted and actual values. It penalizes larger errors more than smaller ones, making it sensitive to outliers. MSE is a widely used metric for regression problems as it provides a clear measure of average prediction error. It is particularly useful when the cost of large errors is high, as it emphasizes larger discrepancies between predicted and actual values.

3) Root Mean Squared Error (RMSE): Root Mean Squared Error is the square root of the Mean Squared Error. It provides an error metric that is in the same units as the target variable and can be interpreted as the standard deviation of the errors. It is sensitive to large errors, making it useful for applications where large discrepancies are particularly problematic.

4) R-Squared (R^2): R-Squared, also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, with higher values indicating better model performance.

MAE is less sensitive to outliers compared to MSE and RMSE. MSE and RMSE penalize larger errors more, providing insight into models' performance in cases where large errors are particularly unfavorable. R^2 provides a normalized measure of model performance, indicating the goodness of fit. Including all these metrics gives a comprehensive view of model performance, addressing different aspects of prediction accuracy and error characteristics.

V. Results And Discussion

The performance of each predictive model based on all four abovementioned evaluation metrics were recorded. The generalizability of the models for predicting biochar yield was investigated by varying the data preprocessing steps. Specifically, the performance was evaluated before and after scaling the features using StandardScaler and handling skewed distributions with logarithmic transformation.

Table 1 presents a comparative analysis of the performance of four models: Linear Regression, Random Forest Regressor, K-Nearest Neighbors (KNN), and Convolutional Neural Networks (CNN) before feature scaling and logarithmic transformation. The Random Forest Regressor outperformed the other models with an MAE of 3.33, MSE of 38.59, RMSE of 6.21, and R^2 of 0.82. However, the overall accuracy of the models was generally lacking, as indicated by the relatively high error metrics and lower R^2 values. This suggests that the raw features might not be well-suited for these models, potentially due to unscaled and skewed data distributions.

Table 2 shows the performance of the same models after feature preprocessing, including scaling and logarithmic transformation. The preprocessing significantly improved the performance metrics across all models. Notably, the KNN model achieved an MAE of 0.08, MSE of 0.03, RMSE of 0.16, and R^2 values of 0.72, indicating a substantial enhancement in prediction accuracy. The Random Forest Regressor also showed improvement, with an MAE of 0.07, MSE of 0.02, RMSE of 0.13, and R^2 of 0.83. These results highlight the importance of proper data preprocessing in improving model performance.

The substantial reduction in error metrics and the increase in R^2 values after preprocessing emphasize the critical role of feature scaling and transformation in model training. The improved results in Table 2 suggest that the models benefit significantly from standardized features and the handling of skewed distributions, leading to more accurate and reliable predictions. This is further supported by the observation from the parity plots shown in Fig. 2-5 which compares the predicted and actual values for two different outputs for the preprocessing scenario of the model.

This study demonstrates that while model selection is important, the preprocessing of input features is equally crucial in achieving optimal model performance, especially in complex regression tasks like predicting biochar yield from food waste pyrolysis.

TABLE I. PREDICTION RESULTS BEFORE FEATURE PREPROCESSIING

Model	Evaluation Metric			
	MAE	*MSE*	*RMSE*	*R^2*
LinearRegression	8.48	106.71	10.33	0.50
RandomForestRegressor	3.33	38.59	6.21	0.82
KNeighborsRegressor	4.59	48.02	6.93	0.78
CNN	5.29	51.95	7.21	0.76

TABLE II. PREDICTION RESULTS AFTER FEATURE PREPROCESSIING

Model	Evaluation Metric			
	MAE	*MSE*	*RMSE*	*R^2*
LinearRegression	0.14	0.03	0.19	0.63
RandomForestRegressor	0.07	0.02	0.13	0.83
KNeighborsRegressor	0.08	0.03	0.16	0.72
CNN	0.08	0.01	0.12	0.85

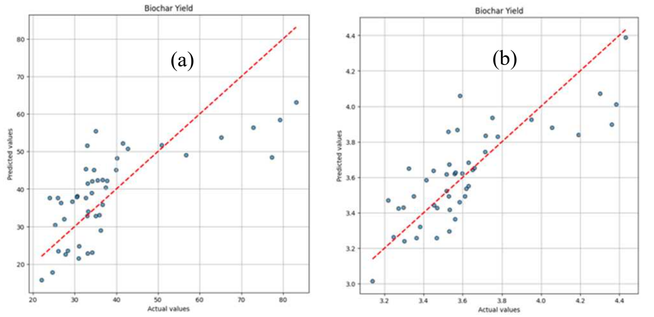

Fig. 2. Parity plots for Linear Regression comparing the actual and predicted values of biochar yield (a) before feature preprocessing (b) after feature preprocessing.

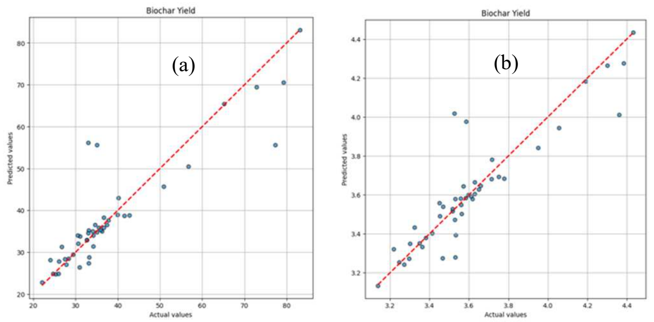

Fig. 3. Parity plots for Random Forest Regressor comparing the actual and predicted values of biochar yield (a) before feature preprocessing (b) after feature preprocessing.

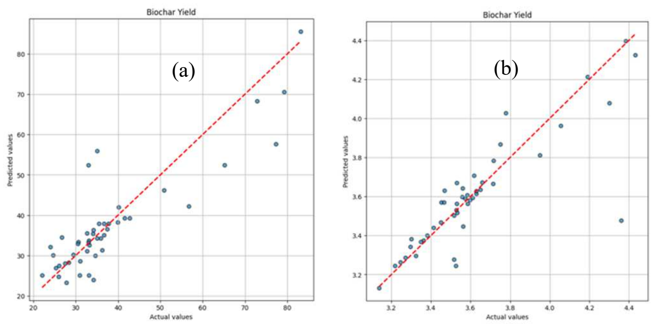

Fig. 4. Parity plots for K-Neighbors Regressor comparing the actual and predicted values of biochar yield (a) before feature preprocessing (b) after feature preprocessing.

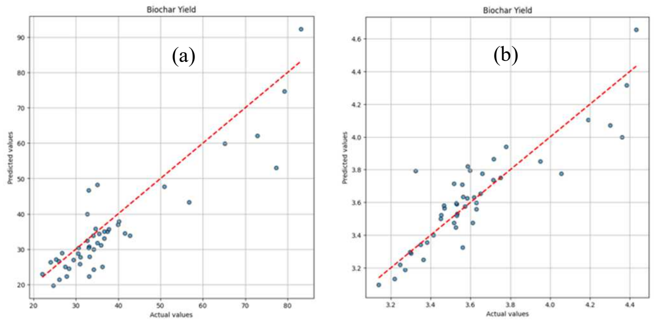

Fig. 5. Parity plots for CNN comparing the actual and predicted values of biochar yield (a) before feature preprocessing (b) after feature preprocessing.

VI. CONCLUSION

In conclusion, this study highlights the potential of AI-driven models in predicting biochar yield from food waste pyrolysis, with a particular focus on Linear Regression, Random Forest, K-Nearest Neighbors, and Convolutional Neural Networks (CNNs). The analysis revealed that data preprocessing, specifically feature scaling and logarithmic transformation, plays a pivotal role in enhancing model accuracy and reliability. The Random Forest Regressor initially demonstrated superior performance among the models, but the overall accuracy improved significantly across all models after preprocessing, emphasizing the importance of data handling in model training.

The results indicate that the KNN and CNN models, in particular, showed remarkable improvement post-preprocessing, achieving lower error metrics and higher R^2 values. This highlights the efficacy of these models in capturing the complex relationships within the dataset. The study also demonstrated that feature standardization and handling skewed data distributions are essential steps for optimizing model performance, as evidenced by the substantial reduction in MAE, MSE, RMSE, and improved R^2 scores.

Ultimately, this research contributes valuable insights into the application of AI algorithms for sustainable biochar production from food waste. The findings recommend for a comprehensive approach combining robust model selection and meticulous data preprocessing to achieve precise and reliable predictions. Our ongoing work is to incorporate explainable AI (XAI) techniques to improve the interpretability of the models. These advancements support the broader goal of sustainable waste management, enhancing resource utilization, and minimizing environmental impact.

ACKNOWLEDGMENT

This research is funded by the Ministry of Higher Education (MOHE) under the Malaysian Technical University Network (MTUN) Strategic Partnership Research Grant. We would also like to thank Universiti Teknikal Malaysia Melaka for supporting this study. Project Number: MTUN/2023/UTEM-FTMK/CRG/MS0003.

REFERENCES

[1] L. Brennan et al., "The role of packaging in fighting food waste: A systematised review of consumer perceptions of packaging," Journal of Cleaner Production, vol. 281, no. 125276, p. 125276, Jan. 2021.

[2] I.-R. Istrate, D. Iribarren, J.-L. Gálvez-Martos, and J. Dufour, "Review of life-cycle environmental consequences of waste-to-energy solutions on the municipal solid waste management system," Resources, Conservation and Recycling, vol. 157, p. 104778, Jun. 2020.

[3] M. F. M. A. Zamri, R. Bahru, F. Suja', A. H. Shamsuddin, S. K. Pramanik, and I. M. R. Fattah, "Treatment strategies for enhancing the removal of endocrine-disrupting chemicals in water and wastewater systems," Journal of Water Process Engineering, vol. 41, p. 102017, Jun. 2021.

[4] Y. Zhang, "Fast microwave-assisted pyrolysis of wastes for biofuels production - A review.", vol. 297, Feb. 2020.

[5] X. J. Lee, H. C. Ong, Y. Y. Gan, W.-H. Chen, and T. M. I. Mahlia, "State of art review on conventional and advanced pyrolysis of macroalgae and microalgae for biochar, bio-oil and bio-syngas production," Energy Conversion and Management, vol. 210, p. 112707, Apr. 2020.

[6] L. Wang et al., "New trends in biochar pyrolysis and modification strategies: feedstock, pyrolysis conditions, sustainability concerns and implications for soil amendment," Soil Use and Management, vol. 36, no. 3, pp. 358–386, May 2020.

[7] Y. Li, R. Gupta, and S. You, "Machine learning assisted prediction of biochar yield and composition via pyrolysis of biomass," Bioresource Technology, vol. 359, p. 127511, Sep. 2022.

[8] Z. Q. S. Nwokediegwu, K. I. Ibekwe, V. I. Ilojianya, E. A. Etukudoh, and O. B. Ayorinde, "Renewable energy technologies in engineering: A review of current developments and future prospects," Engineering Science & Technology Journal, vol. 5, no. 2, pp. 367-384, 2024.

[9] W. A. Indha, N. S. Zamzam, A. Saptari, J. A. Alsayaydeh, and N. bin Hassim, "Development of Security System Using Motion Sensor Powered by RF Energy Harvesting," IEEE Xplore, Sep. 01, 2020.

[10] W. A. Indra, A. W. Y. Khang, Y. T. Tung, and J. A. J. Alsayaydeh, "Radio frequency identification (RFID) item finder using radio frequency energy harvesting," ARPN Journal of Engineering and Applied Science, vol. 14, no. 20, pp. 3554-3560, 2019.

[11] Z. Wang et al., "The role of machine learning to boost the bioenergy and biofuels conversion," Bioresource Technology, vol. 343, p. 126099, Jan. 2022.

[12] A. Hai et al., "Valorization of groundnut shell via pyrolysis: Product distribution, thermodynamic analysis, kinetic estimation, and artificial neural network modeling," Chemosphere, vol. 283, pp. 131162–131162, Nov. 2021.

[13] J. X. Tee, A. Selvarajoo, and S. K. Arumugasamy, "Prediction of carbon sequestration of biochar produced from biomass pyrolysis by artificial neural network," Journal of Environmental Chemical Engineering, vol. 10, no. 3, p. 107640, Jun. 2022.

[14] A. Pathy, S. Meher, and B. P, "Predicting algal biochar yield using eXtreme Gradient Boosting (XGB) algorithm of machine learning methods," Algal Research, vol. 50, p. 102006, Sep. 2020.

[15] M. Khan, Z. Ullah, Ondřej Mašek, Salman Raza Naqvi, and Muhammad, "Artificial neural networks for the prediction of biochar yield: A comparative study of metaheuristic algorithms," Bioresource Technology, vol. 355, pp. 127215–127215, Jul. 2022.

[16] A. Hai et al., "Machine learning models for the prediction of total yield and specific surface area of biochar derived from agricultural biomass by pyrolysis," Environmental technology & innovation, vol. 30, pp. 103071–103071, May 2023.

[17] X. Gong et al., "Short-term Prediction of Multidimensional Data in Power System Based on CNN," Feb. 2023.

[18] K. T. Klasson, "Biochar characterization and a method for estimating biochar quality from proximate analysis results," Biomass and Bioenergy, vol. 96, pp. 50–58, Jan. 2017.

[19] S. S. S. Ranjit, C. F. Tan, and S. K. Subramaniam. "Implementation off-grid solar powered technology to electrify existing bus stop." *2011 Second International Conference on Intelligent Systems, Modelling and Simulation*. IEEE, 2011.

[20] S. A. B. Anas, R. S. S. Singh, and N. A. B. Kamarudin. "Designing an IoT Agriculture Monitoring System for Improving Farmer's Acceptance of Using IoT Technology,". Engineering, Technology & Applied Science Research, vol. 12, no. 1, pp. 8157-8163, 2022.