

# PREDICTIVE MODELLING AND AI-BASED OPTIMIZATION OF INDUSTRIAL BIOETHANOL PRODUCTION: EFFECTS OF FEEDSTOCK YIELD, PROCESS EFFICIENCY AND ENERGY CONSUMPTION

Mr. Balaji Maroti Pulchwad<sup>1\*</sup>, Dr. Abhijit R. Chavan<sup>2</sup>, Dr. Rajkumar Bhimrao Pawar<sup>3</sup>, Dr. Yadnyesh Yashawant Karkare<sup>4</sup>

<sup>1\*</sup>Research Scholar, Chemical Engineering Department, Dr. Babasaheb Ambedkar Technological University, Lonere, Tal. Mangaon Dist. Raigad, 402103 Maharashtra, India. Gmail Id - balajipulchwad1@gmail.com, Orchid ID: 0009 0005 1415 5489

<sup>2</sup>Professor and Head, Chemical Engineering Department, Dr. Babasaheb Ambedkar Technological University, Lonere, Tal. Mangaon Dist. Raigad, 402103 Maharashtra, India.

<sup>3</sup>Assistant Professor, Department of Computer Engineering, Gharda Institute of Technology, A/p Lavel, Tal Khed, Dist: Ratnagiri, 415605, Maharashtra, India. OrchidID: 0000-0002-8910-0117

<sup>4</sup>Lecturer, Chemical Engineering Department, Dr. Babasaheb Ambedkar Technological University, Lonere, Tal. Mangaon Dist. Raigad, 402103 Maharashtra, India.

\*Corresponding Author: Mr. Balaji Maroti Pulchwad, balajipulchwad1@gmail.com

## ABSTRACT

Industrial bioethanol production requires efficient use of feedstock, processing technologies, and energy resources to improve production performance and sustainability. The study investigated the impact of feedstock yield, processing technology performance, and energy usage on industrial bioethanol production and the potential for applying predictive modelling and artificial intelligence for optimization. A secondary data of 2,000 observations and 15 variables related to production, economic, environmental and market were analysed. Production patterns and relationships among major variables were described using descriptive statistics and correlation analysis. Production capacity and bioethanol growth were predicted by Multiple Linear Regression, Random Forest, Gradient Boosting and XGBoost models. The coefficient of determination, mean absolute error and root mean square error were used to evaluate the models, and the feature-importance analysis was used to determine the important predictor variables. Results indicated a weak direct link between feedstock yield, processing efficiency, energy use and production. The importance of the models was comparatively higher for processing technology efficiency, but no single predictor was much more important than the others for production performance. Predictive ability of all predictive models evaluated was poor, with negative test set  $R^2$  values. These results indicate that AI optimization can be a potential exploratory analytical technique, but more process-level information is needed to apply it successfully in industry. Future studies should combine detailed fermentation, pretreatment, feedstock-composition, conversion-yield and real-time operational variables together to enhance predictability and facilitate more effective bioethanol production optimization.

**KEYWORDS:** bioethanol production; artificial intelligence; predictive modelling; process efficiency; energy consumption

## INTRODUCTION

Bioethanol is an important renewable fuel, due to its potential of lowering dependency on fossil fuels, providing a cleaner energy system and assisting the transition towards a more sustainable bioeconomy. Typically, industrial bioethanol processes use biomass-derived raw materials, whose availability, composition and conversion capacity directly impact on the production performance. Hence, increasingly, emphasis is laid on the creation of efficient biorefineries that can convert a variety of biomass resources to fuels and other added value products. Now with the advent of artificial intelligence, modern biorefineries gained an additional boost in their potential through data-driven monitoring, prediction, and optimization of production systems (Abdullah et al., 2025). Lignocellulosic biomass is also of significant importance as it is widely available but needs careful processing to yield desired fuel production, which can be achieved using sustainable methods (Ayub et al., 2025). Simultaneously, AI-driven biomass processing is increasingly being investigated to enhance the efficiency of the conversion process, utilization of resources, and energy integration within industrial biofuel systems (Garg et al., 2025).

However, industrial scale bioethanol production has several issues related to operation and resources. Feedstock yield can vary, impacting the amount of biomass that can be successfully fed into the production process, and the conversion efficiency and production capacity may vary according to the processing technology. Another important consideration is energy use as it can be a significant energy consuming process in pretreatment, fermentation, separation and downstream processing. One solution to improve consistency and minimize inefficiencies throughout biomass production stages is automation of biomass transformation processes (Ananda et al., 2025). Likewise, the inter-connections between intelligent monitoring systems and biofuel production are linked to better process control, higher capacity to respond to the process and more sustainable use of resources (Hussain et al., 2023). Also, prior studies have shown that artificial intelligence can be applied for the prediction of energy-related characteristics of biomass conversion, underscoring the value of data-driven approaches for enhancing the planning of biomass to biofuel processes (Meena et al., 2021).

AI and ML offer a powerful way of analysing complex relationships between production variables that may not be fully captured by traditional methods alone. AI-based techniques can consider simultaneously many operational, environmental and economic variables and detect patterns related to production performances. They have been used in bioenergy systems, moving beyond prediction to optimization, process control, and decision support (Adhikary et al., 2025). AI is now increasingly utilized for modelling and optimization strategies in biorefineries with a data-driven approach, as well as for real-time monitoring (Butean et al., 2025). From prediction, parameter selection, and process improvement, machine learning has emerged as a valuable tool in the smart bioprocessing of complex biological systems (Khanal et al., 2023). Previous studies on AI-driven modelling in bioenergy also highlighted the capability of AI modelling in prediction and optimization of performance in biomass and renewable-energy systems (Liao & Yao, 2021).

AI has a significant impact in fermentation and bioethanol production, where multiple interconnected factors affect the results of the production process. In recent years, the use of AI in fermentation has grown quickly, encompassing tasks like process prediction, monitoring, and control, as well as optimization (Enríquez-García et al., 2025). On a larger scale, AI-driven approaches are increasingly being integrated into the processes of biofuel production, aiding in production planning and decision-making (Kumar & Kuila, 2025). Machine-learning techniques can be also utilized throughout upstream and downstream bioprocess operations; this can make the production systems more data-driven and adaptive (Sharma & Singh, 2025). In the case of bioethanol production, AI is recognized as a crucial asset for process innovation and enhancing the understanding of intricate production data (Srivastava, 2025). The assessment of feedstock yield, process efficiency and energy use in context together in a common predictive framework, however, is important to understand the impact of these integrated factors on industrial bioethanol production. Thus, the present study had prediction modelling and optimisation by AI to analyse factors related to production and determine the conditions that can improve bioethanol performance.

## **OBJECTIVES**

- To evaluate the relationships of feedstock yield, processing efficiency, and energy consumption with industrial bioethanol production performance.
- To develop and compare predictive models for estimating bioethanol production outcomes using operational, economic, and environmental variables.
- To identify the most influential production factors and determine AI-based conditions associated with improved production efficiency and reduced resource use.

## **2. MATERIALS AND METHODS**

### **2.1 Study Design**

The study used a quantitative secondary-data analytical design to explore factors affecting industrial bioethanol production and to create predictive models for production performance. The analysis was primarily based on the feedstock yield, processing technology efficiency and energy requirements, with other economic, environmental, market and operational factors also considered to gain a wider picture about the production situation. Both statistical and machine-learning techniques were combined to extract the relations between the variables and compare the performances of the two methods and determine the variables with significant effect on bioethanol production results.

### **2.2 Data Source and Dataset**

The secondary data set included 2,000 observations of industrial bioethanol production (KhajaAhmed1, 2024). The data comprised 15 variables that accounted for production, feedstock, processing, energy, economic, environmental and market features. These variables included Year, Country, Feedstock Yield, Production Capacity, Processing Technology Efficiency, Energy Consumption, Feedstock Cost, Transportation Cost, Distribution Cost, Carbon Emissions, Water Usage, Market Demand, Price Per Gallon, Government Incentive, and Bioethanol Growth. The values recorded were used as is for analysis.

### **2.3 Study Variables**

The main predictor variables were Feedstock Yield, Processing Technology Efficiency, and Energy Consumption as these variables involved the principal factors included in the study title. Production capacity and Bioethanol growth were considered as important production-performance outcomes. Other predictors were Feedstock Cost, Transportation Cost, Distribution Cost, Carbon Emissions, Water Usage, Market Demand, Price Per Gallon, Government Incentive, Country and Year. These parameters were included to consider the factors that have wider impact on the economic, environmental, temporal and market aspects of industrial bioethanol production.

### **2.4 Data Preparation and Quality Assessment**

The data set was checked for completeness, duplicates, coding consistency, data type and for analytical suitability. Ranges and distributions of numerical variables and consistency of categorical coding were checked. Any missing values and duplicated observations were examined prior to statistical modelling. All continuous variables were kept as numbers and categorical variables were coded as needed for machine learning algorithms. Methods that were sensitive to differences in variable scale were the only ones standardized.

### **2.5 Descriptive and Correlation Analysis**

Data were summarized using descriptive statistics. Mean, standard deviation, minimum and maximum were used to describe continuous variables; and frequency and percentage were used to describe categorical variables. Correlation analysis was used to assess the direction and magnitude of the relationships between key numerical variables. Pearson correlation was used for appropriate continuous variables, and Spearman correlation was used if the assumptions for Pearson correlation were not met. The relationship between production, operational, economic and environmental factors was visualized using a correlation matrix.

## 2.6 Predictive Model Development

Predictive models were developed to estimate the performance of industrial bioethanol production based on the predictor variables present in the data. Conventional baseline approach Multiple Linear Regression and machine-learning models of Random Forest Regression, Gradient Boosting Regression and Extreme Gradient Boosting were performed. The set of predictors comprised Feedstock Yield, Processing Technology Efficiency, Energy Consumption and the other economic, environmental as well as market variables. Based on the predictive analysis, Production Capacity and Bioethanol Growth were taken into account as relevant target variables.

## 2.7 Model Training, Validation and Performance Evaluation

The data set was split into training and testing data sets to allow for the model's performance to be tested on observations not used to develop the model. The predictive models were fitted to the training subset, and the testing subset was used for assessing the generalization performance. For each model, cross validation was used to investigate the model stability if applicable. The coefficient of determination (CD), mean absolute error (MAE), and root mean squared error (RMSE) were used to compare the performance. Increased power to explain and decreased prediction error was taken as an example of increased predictive power.

## 2.8 Feature Importance and AI-Based Optimization

The relative importance of each predictor in the bioethanol production performance was evaluated by feature-importance analysis. Feedstock Yield and Processing Technology Efficiency were considered, in addition to Energy Consumption, and the role of economic, environmental and market-related factors was explored. Finally, the best machine-learning model was used to facilitate the AI-based optimization. It was in the optimization process that the combinations of production conditions associated with improved production performances in the model were detected to achieve greater process efficiency and more efficient energy use.

## 2.9 Statistical Software

All data preparation, statistical analysis, visualization, predictive modelling, model validation, feature-importance assessment and optimization was done with the appropriate statistical and machine-learning software. Data was processed, computed, modelled and presented using python-based data analysis libraries. The analytical workflow was set up to enable the clear comparison of conventional statistical modelling and machine-learning approaches.

## 2.10 Ethical Considerations

This study relied solely on a secondary data set and did not engage in direct recruitment, intervention or interactions with animals or human participants. No participant information in the analysis was from participants who could be personally identified or confidential. So informed consent was not applicable in the present study. The data were used only for research and analysis purposes, and analysis was performed in keeping with the principles of responsible use of data, transparency and research integrity.

## 3. RESULTS

### 3.1 Dataset Characteristics and Production Profile

The final dataset comprised 2,000 observations of industrial bioethanol production characteristics in various countries and years. There were no duplicate records or missing observations found, so all 15 variables could be analysed. Feedstock, processing, production, energy, economic, environmental and market-related factors were represented in the dataset. The descriptive statistics showed significant differences among production capacity, energy use, feedstock attributes, costs, environmental indicators, and bioethanol growth, thus offering a wide foundation for exploring bioethanol production patterns and predictive relationships. Data about the main numerical variables are shown in Table 1.

**Table 1. Descriptive Statistics of Bioethanol Production Variables**

| Variable                         | Mean    | SD     | Minimum | Maximum |
|----------------------------------|---------|--------|---------|---------|
| Feedstock Yield                  | 3.003   | 0.881  | 1.502   | 4.497   |
| Production Capacity              | 175.431 | 43.197 | 100.106 | 249.951 |
| Processing Technology Efficiency | 0.825   | 0.072  | 0.700   | 0.950   |
| Energy Consumption               | 533.732 | 19.579 | 500.029 | 567.970 |
| Feedstock Cost                   | 200.229 | 58.247 | 100.012 | 299.952 |
| Transportation Cost              | 6.524   | 0.855  | 5.007   | 7.997   |
| Distribution Cost                | 14.565  | 2.621  | 10.002  | 18.997  |
| Carbon Emissions                 | 5.468   | 2.580  | 1.000   | 9.999   |
| Water Usage                      | 338.285 | 79.902 | 200.079 | 474.865 |
| Market Demand                    | 137.426 | 21.495 | 100.025 | 174.947 |
| Price Per Gallon                 | 3.010   | 1.147  | 1.006   | 4.994   |
| Bioethanol Growth                | 7.347   | 4.308  | 0.010   | 14.994  |

### 3.2 Relationships Among Feedstock, Process, Energy and Production Variables

The relationships between the main production variables were, in general, not very strong with a correlation analysis. No strong correlation was found between the feedstock yield and production capacity, and a weak positive correlation was found between processing technology efficiency and production capacity. A weak positive relationship was also found between energy use and production capacity. Comparable low correlation was found when bioethanol production was assumed to be the result. These results show that there was no strong linear association between production performance

and any of the three main predictors individually in the current data set. Table 2 displays the correlation coefficients, and the overall trend of relationships is presented in Figure 1.

**Table 2. CORRELATION Matrix of Principal Bioethanol Production Variables**

| Variable            | Feedstock Yield | Process Efficiency | Energy Consumption | Production Capacity | Bioethanol Growth |
|---------------------|-----------------|--------------------|--------------------|---------------------|-------------------|
| Feedstock Yield     | 1.000           | -0.005             | 0.015              | -0.015              | 0.033             |
| Process Efficiency  | -0.005          | 1.000              | -0.015             | 0.038               | -0.055            |
| Energy Consumption  | 0.015           | -0.015             | 1.000              | 0.023               | 0.017             |
| Production Capacity | -0.015          | 0.038              | 0.023              | 1.000               | -0.017            |
| Bioethanol Growth   | 0.033           | -0.055             | 0.017              | -0.017              | 1.000             |

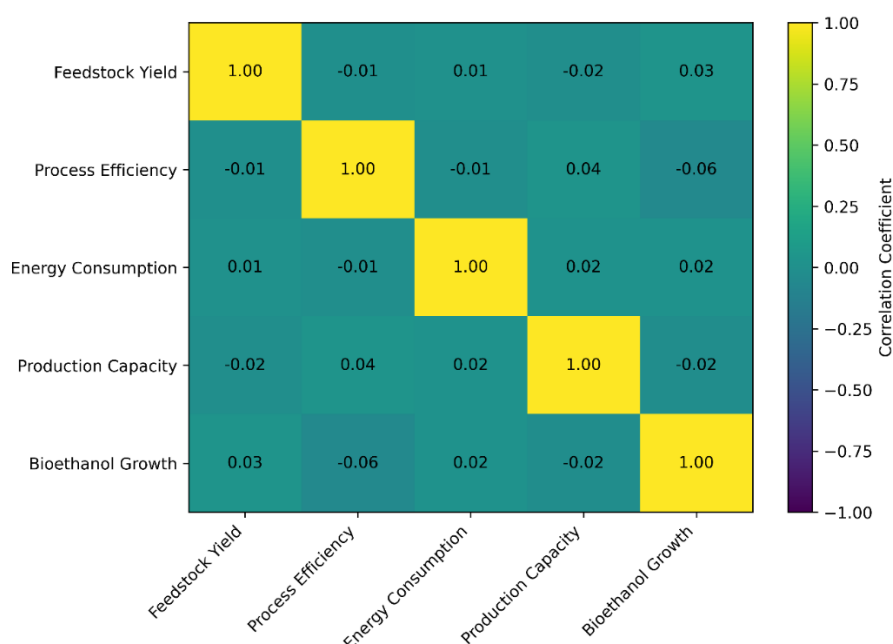


**Figure 1. Correlation Heatmap of Bioethanol Production Variables**

Heatmap showing the direction and magnitude of relationships among the principal operational and production variables.

### 3.3 Country-Level Production Characteristics

At country level, production characteristics were broadly similar across the geographical groups represented. The variations in average feedstock yield, processing efficiency, energy use, production capacity, and bioethanol expansion were relatively small. This indicates that the principal production outcomes did not vary quite significantly due to the geographical grouping. Details of country-specific production characteristics are presented in Table 3.

**Table 3. Country-Level Mean Bioethanol Production Characteristics**

| Country | n   | Feedstock Yield | Process Efficiency | Energy Consumption | Production Capacity | Bioethanol Growth |
|---------|-----|-----------------|--------------------|--------------------|---------------------|-------------------|
| Brazil  | 361 | 3.006           | 0.820              | 534.557            | 176.333             | 7.493             |
| China   | 427 | 2.996           | 0.826              | 533.189            | 176.833             | 7.325             |
| Germany | 430 | 3.089           | 0.827              | 534.179            | 174.961             | 7.351             |
| India   | 404 | 2.920           | 0.823              | 533.826            | 175.483             | 7.142             |
| USA     | 378 | 2.998           | 0.827              | 532.949            | 173.463             | 7.448             |

As shown in Table 3, the major bioethanol production indicators remained broadly distributed across countries without a pronounced separation in overall production performance. Variation in mean production capacity across feedstock yield quartiles is presented in Figure 2. Mean production capacity across increasing levels of processing technology efficiency is shown in Figure 3.

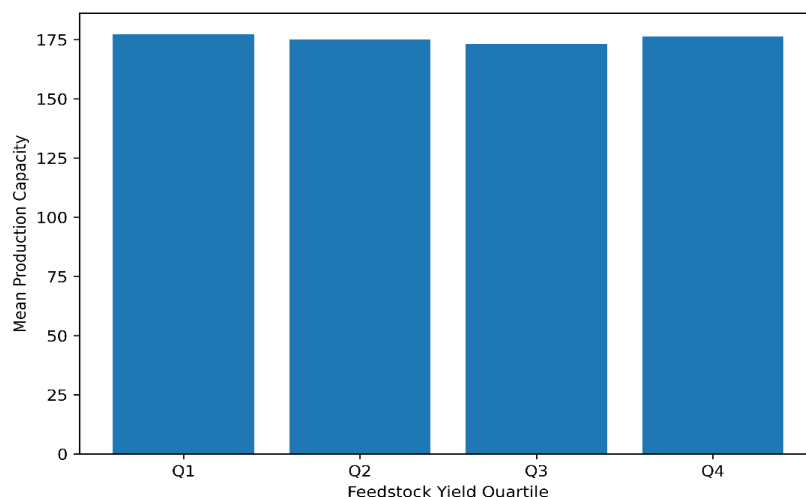


**Figure 2. Mean Production Capacity by Feedstock Yield Quartile**

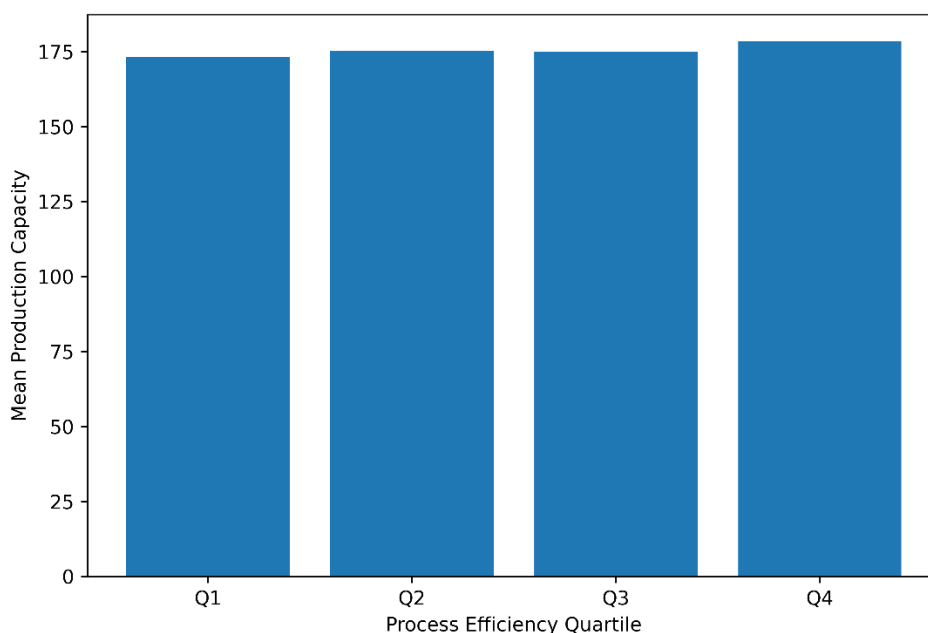


**Figure 3. Mean Production Capacity by Processing Technology Efficiency Quartile**

### 3.4 Predictive Performance of Machine-Learning Models

Four predictive approaches were evaluated for Production Capacity and Bioethanol Growth. Linear Regression, Random Forest, Gradient Boosting and XGBoost were obtained with a small predictive power on the independent test set. Both results had a coefficient of determination under zero, meaning that the models fit the data worse than a baseline model based on the mean value itself. Linear Regression had the lowest prediction error and the least negative  $R^2$  when compared to the methods tested in both target variables. The full model comparison is given in Table 4.

**Table 4. Predictive Performance of Statistical and Machine-Learning Models**

| Target              | Model             | $R^2$  | MAE    | RMSE   |
|---------------------|-------------------|--------|--------|--------|
| Production Capacity | Linear Regression | -0.006 | 35.110 | 41.885 |
| Production Capacity | Random Forest     | -0.022 | 35.343 | 42.209 |
| Production Capacity | Gradient Boosting | -0.043 | 35.706 | 42.636 |
| Production Capacity | XGBoost           | -0.084 | 36.038 | 43.470 |
| Bioethanol Growth   | Linear Regression | -0.024 | 3.510  | 4.116  |
| Bioethanol Growth   | Random Forest     | -0.052 | 3.556  | 4.171  |
| Bioethanol Growth   | Gradient Boosting | -0.083 | 3.604  | 4.232  |
| Bioethanol Growth   | XGBoost           | -0.127 | 3.641  | 4.319  |

### 3.5 Model-Derived Importance of Production Factors

A Random Forest exploratory feature-importance analysis was conducted to investigate the relative contribution of available predictors to Production Capacity. The model revealed the maximum importance for Processing Technology Efficiency followed by Bioethanol Growth, Energy Consumption, Carbon Emissions, and Feedstock Yield. But the relative differences in importance of the leading predictors were relatively small. These importance values should be

viewed as exploratory signals of the model's out-of-sample predictive ability, which is limited, given that the RF model was not a particularly good out of sample predictive model. Table 5 presents the features reported by the leading.

**Table 5. Random Forest Feature Importance for Production Capacity**

| Rank | Predictor                        | Importance |
|------|----------------------------------|------------|
| 1    | Processing Technology Efficiency | 0.091      |
| 2    | Bioethanol Growth                | 0.087      |
| 3    | Energy Consumption               | 0.083      |
| 4    | Carbon Emissions                 | 0.083      |
| 5    | Feedstock Yield                  | 0.082      |
| 6    | Feedstock Cost                   | 0.081      |
| 7    | Transportation Cost              | 0.080      |
| 8    | Price Per Gallon                 | 0.080      |
| 9    | Water Usage                      | 0.079      |
| 10   | Market Demand                    | 0.079      |

The results in Table 5 suggest that production capacity was not dominated by a single predictor and that information was distributed relatively evenly across several operational, environmental, and economic characteristics. The temporal variation in average bioethanol growth across the study years is presented in Figure 4.

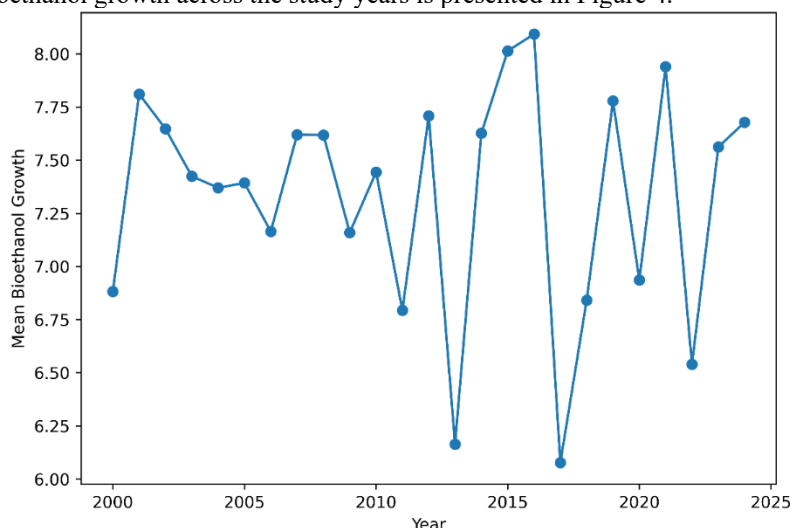


**Figure 4. Average Bioethanol Growth by Year**

### 3.6 Implications for AI-Based Optimization

The out-of-sample accuracy of the predictive analysis was not good enough to allow strong model-based optimization of the industrial production conditions. While the machine-learning models were able to produce fitted predictions and variable-importance estimates, their negative test-set  $R^2$  values suggested that optimized combinations relying solely on the machine-learning models might be unstable or misleading. Hence, the current findings favour an exploratory, AI-assisted decision model instead of optimization suggestions. The dataset is valuable for studying the patterns of production and for testing models of production, but it is not sufficiently comprehensive for robust optimization, which would require including other process-specific variables, stronger correlations between them and the outcome or to be validated using external data from industrial production.

## 4. DISCUSSION

This present study explored the effects of feedstock yield, processing technology efficiency and energy consumption in industrial bioethanol production and its applications to prediction using machine-learning approaches and AI-supported optimization. Generally, the results indicated that the three main variables were not directly related to the production results. There was only a weak linkage between feedstock yield and production capacity and bioethanol growth which indicates that feedstock yield may not be the only factor that can define the production performance. Previous studies also have highlighted that the material properties of the feedstocks should be evaluated in conjunction with the conversion technologies, pretreatment conditions and other production factors (Hasan et al., 2025; Narasimman et al., 2025).

The relative importance of the model-derived efficiency of the processing technology was slightly higher than the other leading factors but was still low in direct correlation with production capacity. This indicates the possibility of the process efficiency having a combined or nonlinear effect on production. AI-driven solutions have been increasingly adopted to detect such intricate production patterns and optimize conversion performance (Damian et al., 2024). Efficiency in enzymatic and biomass processing could also positively affect the production of biofuel even at the low level of single variable effects (Saxena, 2025). Other predictive analyses have emphasised the importance of considering several

processes features together when predicting the conversion of biomass into biofuels (Mateescu et al., 2022; Priya & Yadav, 2025).

There was also a weak correlation between energy consumption and production capacity. However, there was no direct relationship between increased energy consumption and increased production efficiency. This result is an indication of the necessity to consider energy use along with resource use and environmental impacts. AI and machine-learning methods have been applied in the field of biofuels to optimize the use of resources and promote production sustainability (Mafat et al., 2024), and recent studies focusing on biomass valorization highlight the need to consider the conversion efficiency, energy consumption, and environmental footprint of the process (Saleh et al., 2025).

The out-of-sample performance of the predictive models was limited. All the models (Linear Regression, Random Forest, Gradient Boosting, and XGBoost) gave negative  $R^2$  values of the test-set, suggesting that the available predictors were not enough to reliably predict production capacity and bioethanol growth. While machine learning has demonstrated promising potential for biofuel applications, data quality, variable selection, and the availability of process-specific information play a crucial role in the performance of these models (Ajao et al., 2024). This is consistent with microalgal and biodiesel modelling studies which reported different levels of predictive performance depending on the structure of data sets and model inputs (Imamoglu, 2024; Xing et al., 2021).

The results of the feature-importance analysis showed that the processing efficiency, bioethanol growth, energy consumption, carbon emissions, and feedstock yield were all roughly equal in the amount of predictive information they provided. It is important to therefore consider these findings to be exploratory, not causal. Past works on AI-optimized biofuel production have highlighted the importance of combining several production factors to determine the best conditions (Budi Susilo et al., 2025, Nagappan et al., 2025, Zakir Hossain et al., 2022).

The results show that, from an industrial point of view, it would be beneficial to have more detailed plant-level variables such as the fermentation conditions, the parameters used for pretreatments, the composition of the feedstock, and conversion yields for AI-based optimization. Additionally, the economic viability must be addressed as the cost of implementation can affect the practicality of AI-assisted biofuel production systems. To achieve greater reliability in prediction and use in industry, more detailed process-level and externally validated data should therefore be used in future studies.

## 5. CONCLUSION

This study analysed the interdependencies of feedstock yield, processing technology efficiency and energy consumption on industrial bioethanol production and determined the applicability of statistical and machine-learning models for prediction and AI-driven optimization. The results indicated that the three major production factors had weak direct links to bioethanol development and production capacity. The model-derived importance of processing technology efficiency was relatively high, with no single factor being clearly most important in production performance. Energy use and feedstock production were also factors in the overall production profile but had only a limited independent influence within the current data set. Among the four models evaluated (Linear Regression, Random Forest, Gradient Boosting, and XGBoost), none of the models performed well in terms of out-of-sample predictive performance. This suggests that the production variables provided is inadequate to predict production results and that no algorithm was able to compensate for the limitations of the variables. However, feature-importance analysis indicated that production performance depended on a variety of operational, environmental, and economic factors, none of which was the single determining factor. In conclusion, the study highlights the potential of AI-assisted analysis for exploring more industrial bioethanol systems and underscores the significance of data quality and process-specific information. More detailed fermentation, pretreatment, feedstock-composition, conversion-yield and real-time plant variables should be accounted for in future predictive and optimization frameworks to enhance predictive accuracy and facilitate more reliable industrial decision-making.

## REFERENCES

1. Abdullah, M., Malik, H. A., Ali, A., Boopathy, R., Vo, P. H., Danaee, S., ... & Malik, S. (2025). AI-driven algae biorefineries: A new era for sustainable bioeconomy. *Current Pollution Reports*, 11(1), 21.
2. Adhikary, S., Sau, S., Debnath, S., & Ghosh, D. (2025). Advancing bioenergy: leveraging artificial intelligence for efficient production and optimization. In *Computer Vision and Machine Intelligence for Renewable Energy Systems* (pp. 177-186). Elsevier.
3. Ajao, J., Akanbi, O., Akinwale, I., Adedokun, A., Ogunsakin, O., & Ige, A. (2024). Research Advances on Machine Learning Technologies for Enhanced Biodiesel Production: A Comprehensive Review. *Petroleum & Coal*, 66(1), 1.
4. Ananda, A., Sujeeth, R. K., & Archana, S. (2025). Integrating automation in biomass transformation: Opportunities, challenges, and future directions. *BioEnergy Research*, 18(1), 63.
5. Ayub, A., Singh, R., Wani, A. K., Xie, R., Ahmad, F., Gorish, B. M. T., ... & Dar, M. A. (2025). Role of artificial intelligence (AI) in sustainable biofuel production from lignocellulosic biomass. In *Artificial intelligence applications for a sustainable environment* (pp. 329-348). Cham: Springer Nature Switzerland.
6. Budi Susilo, Y. K., Loganathan, K., Sukirthanandan, P., & Abdul Rahman, S. (2025). Sustainable Biofuel From Algae: An Artificial Intelligence Optimization Approach for Production Enhancement. *Sustainable Biofuel from Algae: An Artificial Intelligence Optimization Approach for Production Enhancement* (April 29, 2025).
7. Butean, A., Cutean, I., Barbero, R., Enriquez, J., & Matei, A. (2025). A review of artificial intelligence applications for biorefineries and bioprocessing: from data-driven processes to optimization strategies and real-time control. *Processes*, 13(8), 2544.

8. Damian, C. S., Devarajan, Y., Thandavamoorthy, R., & Jayabal, R. (2024). Harnessing artificial intelligence for enhanced bioethanol productions: a cutting-edge approach towards sustainable energy solution. *International Journal of Chemical Reactor Engineering*, 22(7), 719-727.
9. Enríquez-García, H. C., Salcedo-Medina, F. D. J., & Mateos-Díaz, J. C. (2025). Artificial intelligence and fermentation: applications, publications and trend analysis. *Revista de ciencias tecnológicas*, 8(4).
10. Garg, R., Rajput, P., Vibhandik, J., Ali, A., & Abrar, I. (2025). Advances in AI-driven biomass processing: A review of conversion technologies, optimization strategies, and smart energy integration. *ACS omega*, 10(42), 49300.
11. Hasan, M. S., Dalim, H. M., & Chowdhury, B. R. (2025). AI-Driven Forecasting of US Biofuel Production and Feedstock Demand: A Machine Learning Analysis. *Journal of Environmental & Agricultural Studies*, 6(4), 22.
12. Hussain, Z., Babe, M., Saravanan, S., Srimathy, G., Roopa, H., & Boopathi, S. (2023). Optimizing biomass-to-biofuel conversion: IoT and AI integration for enhanced efficiency and sustainability. In *Circular economy implementation for sustainability in the built environment* (pp. 191-214). IGI Global Scientific Publishing.
13. Imamoglu, E. (2024). Artificial intelligence and/or machine learning algorithms in microalgae bioprocesses. *Bioengineering*, 11(11), 1143.
14. KhajaAhmed1. (2024). Bio ethanol production dataset [Data set]. Kaggle. <https://www.kaggle.com/datasets/khajaahmed1/bio-ethanol-1>.
15. Khanal, S. K., Tarafdar, A., & You, S. (2023). Artificial intelligence and machine learning for smart bioprocesses. *Bioresource Technology*, 375, 128826.
16. Kuila, A., & Kumar, D. (Eds.). (2025). *Optimizing biofuel production with artificial intelligence*. John Wiley & Sons.
17. Liao, M., & Yao, Y. (2021). Applications of artificial intelligence-based modeling for bioenergy systems: A review. *GCB Bioenergy*, 13(5), 774-802.
18. Mafat, I. H., Palla, S., & Surya, D. V. (2024). Machine learning and artificial intelligence for algal cultivation, harvesting techniques, wastewater treatment, nutrient recovery, and biofuel production and optimization. In *Value added products from bioalgae based biorefineries: opportunities and challenges* (pp. 463-487). Singapore: Springer Nature Singapore.
19. Mateescu, C., Tudor, E., Dima, A. D., Chiriță, I., Tanasiev, V., & Prisecaru, T. (2022, May). Artificial intelligence approach in predicting biomass-to-biofuels conversion performances. In *2022 23rd International Carpathian Control Conference (ICCC)* (pp. 370-375). IEEE.
20. Meena, M., Shubham, S., Paritosh, K., Pareek, N., & Vivekanand, V. (2021). Production of biofuels from biomass: Predicting the energy employing artificial intelligence modelling. *Bioresource Technology*, 340, 125642.
21. Nagappan, B., Subbiah, G., Paliwal, R. K., Choudhury, S., Rai, K., Mehar, K., ... & Kamakshi Priya, K. (2025). Data-driven optimization of biomass conversion pathways: integrating thermochemical processes. *International Journal of Chemical Reactor Engineering*, 23(11), 1309-1326.
22. Narasimman, K., Pillai, A., Anand, A., Pillai, A., & Mathimani, T. (2025). Maximizing sustainable bioethanol production from food waste: progress in diverse pretreatment methods and integration of artificial intelligence. *Sustainable Energy & Fuels*, 9(22), 6097-6115.
23. Priya, R., & Yadav, P. (2025). Cost Analysis of Artificial Intelligent-Based Biofuels Industry. *Optimizing Biofuel Production with Artificial Intelligence*, 285-310.
24. Saleh, Y., Ali, L., & Altarawneh, M. (2025). Recent advances in biomass valorization through thermochemical processes, bio-oil production and AI strategies: a concise review. *RSC advances*, 15(54), 45943-45978.
25. Saxena, R. R. (2025). *Bioenergy Production Enhancement and Optimization Through Smart Enzyme Engineering*.
26. Sharma, D., & Singh, K. (2025). AI-enhanced bioprocess technologies: machine learning implementations from upstream to downstream operations. *World Journal of Microbiology and Biotechnology*, 41(8), 278.
27. Srivastava, A. (2025). Revolutionizing bioethanol production: The role of AI in process innovation. In *Methods in Microbiology* (Vol. 56, pp. 167-190). Academic Press.
28. Xing, Y., Zheng, Z., Sun, Y., & Agha Alikhani, M. (2021). A review on machine learning application in biodiesel production studies. *International Journal of Chemical Engineering*, 2021(1), 2154258.
29. Zakir Hossain, S. M., Sultana, N., Irfan, M. F., Haque, S. M., Nasr, N., & Razzak, S. A. (2022). Artificial intelligence-based super learner approach for prediction and optimization of biodiesel synthesis—A case of waste utilization. *International Journal of Energy Research*, 46(14), 20519-20534.