



## Journal of Business Intelligence and Data Analytics

Journal homepage: [www.sciforce.org](http://www.sciforce.org)

# Predictive Modeling of Process Parameters in WCO-Based Biodiesel Production Using Advanced Regression Techniques

Tejasvi Gorre\*

*Home Depot Management Company LLC, 2455 Paces Ferry Rd SE, Atlanta, USA*

## ARTICLE INFO

## Article history:

Received: 20250312

Received in revised form: 20250318

Accepted: 20250325

Available online: 20250329

*Keywords:*

Waste Cooking Oil (WCO),  
Biodiesel Production,  
Machine Learning Regression,  
Process Parameter Optimization,  
Sustainable Biofuel and Predictive  
Modeling.

## ABSTRACT

Biodiesel production from waste cooking oil (WCO) presents a compelling opportunity to transform discarded oil into a renewable energy resource. Through the conversion of WCO into biodiesel, not only is waste effectively reduced, but a greener, more sustainable alternative to conventional fossil fuels is provided—furthering the shift towards environmentally conscious energy solutions. The importance of this research cannot be overstated. It plays a crucial role in advancing sustainable energy practices, especially by tapping into WCO as a viable and underutilised feedstock for biodiesel production. Consider the scale of global WCO generation: in Canada alone, 135,000 tons are produced annually, while in Asia, the figures soar to a staggering 5.5 million tons. The vast potential for converting this surplus waste into high-value biofuel not only promises substantial environmental benefits but also unlocks significant economic opportunities. The methodology leveraged three distinct machine learning models: Linear Regression (LR), Random Forest Regression (RFR), and Support Vector Regression (SVR).

These models were rigorously trained and tested on experimental data derived from biodiesel production processes. The study delved into four critical parameters: Free Fatty Acid (FFA) content, fluctuating between 1.7% and 3.5%, moisture percentage ranging from 0.05% to 0.3%, viscosity measured at 35 to 43 cSt, and reaction time spanning 2 to 3.3 hours. The results were striking, underscoring the robust predictive power of all three models. SVR stood out, achieving the highest training accuracy ( $R^2 = 0.998$ ), while RFR exhibited a remarkable ability to generalise well on unseen test data ( $R^2 = 0.989$ ). The analysis uncovered compelling correlations: notably, a robust negative relationship between FFA content and biodiesel yield (-0.91), alongside a positive correlation between viscosity and yield (0.85).

These findings underline the capacity of machine learning models to accurately predict biodiesel yields from waste cooking oil (WCO). Each model revealed unique strengths, yet even the simpler Linear Regression model, with an impressive  $R^2$  of 0.979 on test data, pointed to a predominantly linear link between the process parameters and the final yield. Such insights provide invaluable guidance for refining industrial biodiesel production processes, championing the shift towards sustainable energy alternatives and addressing the pressing issues of waste management.

© Tejasvi Gorre.

\*Corresponding author. e-mail: [tejasvigr@gmail.com](mailto:tejasvigr@gmail.com)

## **Introduction.**

The relentless expansion of the global population is driving an astonishing surge in energy demands, with projections indicating a striking 53% increase by 2030 compared to figures from 2001. This growing demand is unfolding hand in hand with the rapid exhaustion of non-renewable fossil fuels—coal, oil, and gas—resources that are, at best, expected to last for only another 200, 40, and 70 years, respectively, based on current consumption patterns. However, as fossil fuels continue to fuel essential industries like transportation, manufacturing, and electricity generation, the environmental repercussions intensify, with escalating carbon emissions amplifying the overarching issue of global climate change [1]. In the face of these mounting concerns, the need to explore alternative, more sustainable energy sources has never been more pressing. Solar, wind, nuclear, hydro, and biofuels have all emerged as promising contenders.

Among them, biofuels—particularly biodiesel—stand out as renewable energy sources capable of curbing carbon emissions. When derived from non-edible feedstock's, biodiesel offers a viable alternative to conventional diesel fuel, marking a crucial step towards cleaner energy solutions [2]. The surge in fossil fuel consumption, driven by the forces of economic globalization, the rapid expansion of the global population, and the relentless march of industrialization, has significantly contributed to the rise in greenhouse gas emissions. This, in turn, has led to an alarming accumulation of carbon in the atmosphere, with far-reaching consequences for the global climate. In an effort to address these escalating challenges and enhance energy security, nations across the world are increasingly diversifying their energy portfolios. Among these alternatives, biofuels—hailing from sustainable sources with the right chemical properties—have emerged as a viable substitute for traditional fossil fuels [3].

However, the rise of first-generation biofuels, which are derived from edible oils, has sparked concerns over food security. This concern has pushed the energy sector to seek out alternative, less contentious fuel sources. Enter the second generation of biofuels, which focuses on utilizing residual biomass and waste, such as waste cooking oil (WCO). WCO, a by-product of cooking processes, is rich in free fatty acids, making it a promising and technically feasible feedstock for biofuel production. With its abundance and cost-effectiveness, sourced from a wide range of establishments—including restaurants, food processing industries, fast-food chains, and households—WCO stands out as an attractive alternative [4]. Waste cooking oil (WCO), an often overlooked by-product sourced from households, restaurants, hotels, and food processing industries, has carved out a notable niche in global biodiesel production, contributing around 10% to the total output. Abundant and low-cost, WCO seems like a promising

resource; however, its conversion into biodiesel is far from straightforward.

The high levels of free fatty acids present in WCO complicate the process, rendering it economically challenging. On top of this, technical barriers—such as shortages of raw feedstock and the complexities surrounding collection logistics—hinder its broader use. In light of these issues, the recycling of WCO into renewable resources, coupled with ongoing advancements in conversion technology, is of paramount importance [5]. Figures from multiple countries highlight the vast scale at which WCO is produced, underscoring the magnitude of the problem. In Canada alone, approximately 135,000 tons of WCO are generated annually. Meanwhile, the UK and various European Union nations see figures ranging from 200,000 to 1,000,000 tons per year. But perhaps most strikingly, Asia grapples with the production of roughly 5.5 million tons of WCO each year. In Thailand, the situation is particularly dire, with 117,000 tons disposed of annually without adequate treatment. These staggering volumes of WCO raise significant concerns surrounding the logistics of collection, treatment, and disposal, highlighting the urgent need for sustainable and effective solutions [6].

Recent investigations have ventured deep into the use of a variety of ash materials—peanut shell ash, coal fly ash, and even banana peel ash—leveraging their catalytic properties to foster biodiesel production. These materials not only exhibit catalytic activity but also offer a sustainable approach to waste repurposing. Building on earlier research, this study shifts focus to explore the untapped potential of wheat shell ash and water scale, harnessing them as rich sources of calcium oxide for biodiesel synthesis. The aim is to optimize reaction conditions, employing methanol and waste cooking oil (WCO) as the reactants, in pursuit of enhanced efficiency [7]. Biofuels, ranging from biodiesel to bioethanol, emerge as crucial contenders in the battle to mitigate carbon emissions within the transport sector. Among them, biodiesel stands out—especially as a replacement for conventional petroleum diesel. In fact, biodiesel accounts for nearly 80% of the biofuel production across the EU, underscoring its growing prominence.

Furthermore, biodiesel derived from used cooking oil (UCO) or waste cooking oil (WCO) is classified as a second-generation biofuel, distinguished by its non-crop feedstock base. Not only does this biofuel offer remarkable potential in terms of quality, but it also promises a significant reduction in production costs [8]. In countries such as Greece, where the consumption of vegetable oil is exceptionally high, the surplus of used cooking oil (UCO) has become a pressing issue. The conversion of this UCO into biodiesel presents a forward-thinking and sustainable solution, addressing both waste management and energy demands.

However, the careless disposal of UCO into sewage systems brings with it a host of environmental and economic challenges,

ranging from water contamination to increased operational costs for wastewater treatment facilities. Recycling UCO into biodiesel emerges as a robust remedy, offering a way to not only reduce harmful environmental impacts but also enhance energy security [9]. In India, one of the largest consumers of cooking oil globally, vast quantities of waste cooking oil are produced each year. Converting this waste into biodiesel tackles the dual problems of waste disposal and energy generation, while promoting the concept of "waste to wealth." This shift toward intelligent waste management practices has the potential to transform the nation's waste into a valuable resource. With strategic initiatives in place, India is well-positioned to harness a significant portion of its used cooking oil for biodiesel production, strengthening its energy security in the process [10].

The economic feasibility of biodiesel production can be markedly enhanced through the adoption of efficient heterogeneous catalysts, which offer distinct advantages over their homogeneous counterparts, notably in terms of recyclability and a reduced environmental footprint. A growing body of research has underscored the potential of these catalysts, such as barium oxide (BaO) supported on various substrates, in facilitating Transesterification reactions for biodiesel synthesis. In addition, catalysts built upon tin oxide (SnO<sub>2</sub>) have showcased promising catalytic activity, positioning them as excellent candidates for producing biodiesel from waste cooking oil [11]. Meanwhile, non-edible oils, including jatropha oil, castor oil, and waste cooking oil (WCO), are gaining increasing traction as viable feedstocks for biodiesel production.

Among them, WCO stands out, thanks to its cost-effectiveness, the avoidance of competition with edible oils, and its potential to address disposal issues. Life Cycle Assessment (LCA) studies have consistently highlighted the environmental advantages of utilizing WCO for biodiesel, prompting a surge in research aimed at its economic viability [12]. To truly gauge the economic feasibility of biodiesel derived from WCO, a thorough analysis of factors like catalyst costs and process optimization is essential. Despite the dominance of homogeneous catalysts, there is a growing interest in CaO-based catalysts, driven by their simplicity and outstanding catalytic performance. Tools like Aspen Plus have become indispensable in refining the biodiesel production process, helping to evaluate both its technical and economic feasibility [13]. Response Surface Methodology (RSM) has, over time, emerged as a cornerstone in the meticulous optimization of biodiesel production parameters.

Its application facilitates not only the maximization of yield but also ensures the minimization of operating costs—an invaluable dual benefit. Numerous studies have underlined the prowess of RSM in enhancing process parameters, all while ensuring that the biodiesel produced adheres to stringent fuel quality standards. The exploration of the kinetics and thermodynamics involved in biodiesel production, particularly from Waste Cooking Oil (WCO), has been the subject of

extensive research, revealing its potential to significantly reduce production expenses when compared to alternative feedstock's. What's more, the fusion of RSM with the desirability function approach introduces an innovative methodology that not only optimises production but also delves into the complexities of reaction mechanisms and system dynamics [14]. When considering waste biomass, it becomes evident that diverse sources—ranging from agricultural by-products and sewage to mining residues, including those from the iron and steel industry—offer an abundance of material for catalyst production. Take agricultural waste, like rice straw, for example. Not only does its utilization for catalyst synthesis breathe new life into what would otherwise be discarded, but it also alleviates the environmental burden linked to their disposal. In parallel, there is a discernible shift toward solid acid catalysts derived from carbon-based materials, such as sulfonated cellulose or glucose. This marks a clear preference for catalysts that are both reusable and more environmentally friendly, thus positioning them as viable alternatives to traditional liquid acids [15]. In biodiesel production, the choice of catalyst hinges on the free fatty acid (FFA) content of the feedstock, which dictates the entire process.

Alkaline catalysts prove effective when the FFA content is low, but as FFA levels rise, their use becomes problematic, necessitating alternative catalysts to prevent the dreaded saponification reaction. While enzyme catalysts offer a green, non-polluting alternative, their high cost often renders them an impractical choice in many cases. Enter concentrated sulfuric acid, a versatile agent capable of catalyzing both esterification and transesterification processes. However, this solution is not without its drawbacks, such as the corrosion of equipment and the generation of toxic wastewater, highlighting the pressing need for heterogeneous acid catalysts to offer a more sustainable solution [16]. In a shift towards sustainability, waste cooking oil (WCO) has emerged as an economically viable feedstock for biodiesel production, providing an ingenious solution to the environmental hazards of improper disposal. By converting WCO into biodiesel, environmental pollution is curbed, offering a dual benefit: protection for human health and aquatic ecosystems, alongside a significant reduction in the costs tied to waste management [17].

## MATERIALS AND METHODS

### Linear Regression:

Linear regression, a cornerstone in both statistical and machine learning realms, serves as a powerful tool to unravel the intricate relationships between independent variables and a dependent outcome. Its utility stretches far beyond simplicity, thanks to its adaptability and interpretability, cementing its place as a cornerstone in predictive modeling and data analysis. The core aim of this method is elegantly straightforward: to uncover the linear equation that best captures the underlying patterns in data by minimizing the residual errors [18].

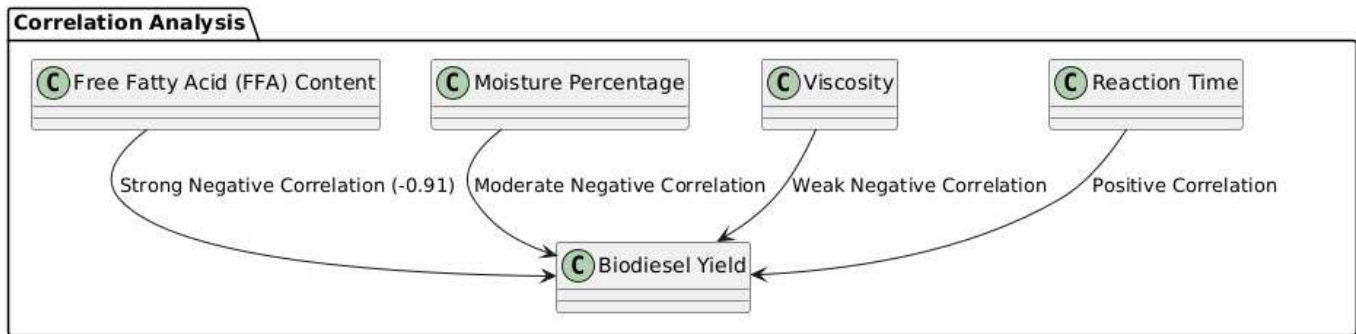
The linear regression model can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

In this equation,  $y$  represents the predicted output, while  $x_1, x_2, \dots, x_n$  denote the input features—these could range from process parameters such as FFA content, viscosity, moisture percentage, and more. The intercept  $\beta_0$  anchors the equation, and  $\beta_1, \beta_2, \dots, \beta_n$  are the coefficients, dynamically estimated during the training phase. Lastly,  $\epsilon$  accounts for the residual error, capturing the discrepancy between the model's prediction and the actual observed data.

Linear regression finds its application across a diverse range of fields, each leveraging its power to extract meaningful insights from complex data. In economics, for instance, it's routinely employed to predict crucial indicators such as GDP growth, inflation, and unemployment. Take, for example, the use of linear regression to forecast GDP, with macroeconomic

factors playing a pivotal role in shaping the predictions. In healthcare, this method has proven indispensable, enabling the estimation of patient recovery times by accounting for clinical variables like age and disease severity [19-20]. But the scope doesn't end there. In environmental science, linear regression comes to the forefront when modeling carbon dioxide emissions, linking them intricately to energy consumption and industrial output. Meanwhile, in the realm of marketing, the technique is put to work predicting product sales, factoring in variables like advertising expenditure and market conditions. In education analytics, too, it holds great promise, with linear regression used to explore the correlation between student performance, study hours, and attendance. Finally, in the competitive world of real estate, linear regression offers a robust tool for estimating property prices, considering elements like size, location, and amenities [21]. Linear regression stands out primarily for its simplicity and ease of interpretation, making it a go-to tool in various analytical settings. Yet, lurking beneath its straightforwardness are its assumptions—linearity, independence, homoscedasticity, and the normality of residuals—which can confine its scope. When these



foundational assumptions crumble under the weight of real-world data, alternative methods such as polynomial regression or careful data transformations often become necessary. Still, despite its limitations, linear regression is far from obsolete. Its enduring value lies in its ability to distil intricate relationships

into clear, comprehensible models. This remarkable feature ensures its continued relevance in research and decision-making, even as the landscape of computational techniques evolves [22].

### Random Forest Regression:

Random Forest Regression stands as a powerful ensemble learning algorithm, meticulously crafted to enhance prediction accuracy while curbing the often-dreaded issue of overfitting—particularly when grappling with intricate datasets. This method constructs a multitude of decision trees during its training phase, then aggregates their individual outputs to offer a prediction that is both robust and precise. Through this averaging technique, it deftly reduces variance, positioning itself as the go-to method for handling non-linear regression challenges [23].

At its core, Random Forest Regression relies on a collective of decision trees, each contributing to the final output. The prediction formula is expressed as:

$$y = \frac{1}{N} \sum_{i=1}^N f_i(x)$$

Here,  $y$  denotes the predicted outcome, while  $N$  indicates the total number of trees in the forest. The term  $f_i(x)$  reflects the prediction made by the  $i$ -th tree, each offering a distinct estimate based on its respective parameters.



A decision tree works by partitioning data into branches based on the values of various features, recursively dividing the dataset to make accurate predictions. The decision-making process behind each split is rooted in the objective of minimizing loss functions—commonly the Mean Squared Error (MSE) in regression problems. However, while effective; single decision trees tend to struggle with overfitting and exhibit high variance, particularly when tasked with complex datasets. This is where Random Forests step in, addressing these limitations by aggregating multiple decision trees into a powerful ensemble model [24]. To boost accuracy and combat overfitting, Random Forest employs two key strategies.

The first is Bootstrap Aggregation, or Bagging, wherein each tree is trained on a randomly chosen subset of the data. The second, Feature Randomization, ensures diversity by selecting a random subset of features at each node split. Together, these techniques foster a variety of decision trees, significantly enhancing the model's ability to generalize and curbing overfitting [25]. Random Forest excels at managing both numerical and categorical data with remarkable efficiency. It doesn't just handle these data types; it goes a step further, automatically imputing missing values while delivering insightful feature importance rankings. This algorithm's true versatility emerges from its scalability, comfortably handling high-dimensional datasets, all while maintaining robustness in the face of noisy data. It's a tool that proves indispensable across a multitude of applications [26].

The performance of this model hinges significantly on certain hyper parameters: the number of trees ( $n\_estimators$ ), the maximum depth of each tree ( $max\_depth$ ), and the minimum samples required for a split ( $min\_samples\_split$ ). These elements don't merely influence the outcome—they can make or break the efficiency of the model. To uncover the optimal combination, techniques such as grid search and cross-validation are employed, fine-tuning the model to perfection and ensuring the best possible configuration for success [27]. Random Forest finds its application in diverse fields, from finance, where it's pivotal in risk modeling, to healthcare, where it aids in disease prediction, and even agriculture, where it plays a crucial role in forecasting yields.

In contrast to linear regression, it excels in uncovering non-linear relationships that would otherwise go unnoticed. While Gradient Boosting may, on occasion, edge ahead in accuracy, its need for meticulous tuning makes it a more delicate tool to wield [28]. The strength of Random Forest Regression lies not only in its ability to model intricate, non-linear relationships but also in its inherent robustness and scalability. This combination allows it to handle vast, diverse datasets with ease, making it a formidable force in predictive analytics. Its broad-reaching applications across various sectors underscore its significance in the ever-evolving landscape of modern machine learning.

### Support Vector Regression (SVR):

Support Vector Regression (SVR) stands as a formidable statistical technique, stretching the boundaries of traditional regression by borrowing principles from Support Vector Machines (SVM). What sets SVR apart from conventional regression methods is its ability to focus on finding a function that deviates from actual observed values only by a margin no greater than a pre-defined threshold. This inherent trait provides SVR with an unparalleled robustness, especially when confronted with outliers, while simultaneously offering a remarkable degree of flexibility in capturing complex, often nonlinear, relationships [29]. At the heart of SVR lies the transformative power of kernel functions. These mathematical tools map the input data into high-dimensional feature spaces, enabling the detection of intricate, non-linear patterns that are often hidden in the raw data. Through this transformation, SVR does not just attempt to fit a line but uncovers relationships that might elude simpler, linear models. The choice of kernel is, however, a critical factor—the kernel function dictates the geometry of the data's transformation and, in turn, heavily influences the overall performance and accuracy of the model [30].

The prediction in SVR follows a succinct yet profound equation, reliant on a chosen kernel function:

$$y = \sum_{i=1}^n \alpha_i K(x_i, x) + b$$

Here,  $y$  denotes the predicted output,  $\alpha_i$  represents the Lagrange multipliers,  $K(x_i, x)$  is the kernel function (e.g., RBF or polynomial kernel),  $x$  is the input vector (which includes process parameters), and  $b$  is the bias term.

A variety of kernels—linear, polynomial, and radial basis function (RBF)—are commonly employed in Support Vector Regression (SVR), each one tailored to different datasets and the complexities inherent in their relationships. Upon choosing the appropriate kernel, SVR seeks to minimize error, but not at the cost of over-complicating the model; it strives to maintain a delicate balance between bias and variance, ensuring optimal performance [31]. This balance is achieved through a tolerance margin, permitting some error, yet ensuring the model remains sufficiently close to the data to capture its underlying trends, without succumbing to overfitting. Far from being a mere technicality, this feature significantly boosts the model's predictive prowess, while also shielding it from the disruptive influence of noise and outliers in the dataset [32].

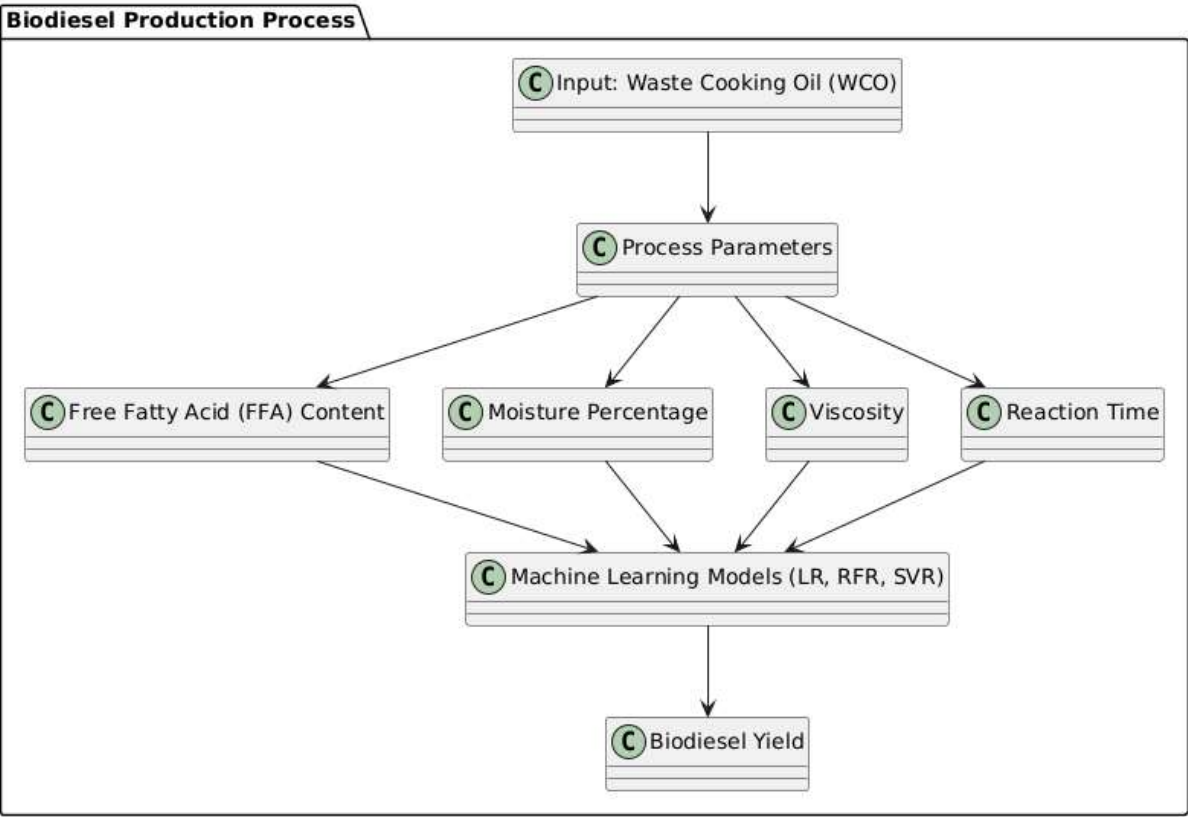
Moreover, the true potential of SVR is realized only when its hyper parameters are finely tuned—parameters like the regularization term and those specific to the chosen kernel. Adjusting these allows for a model more attuned to the nuances of particular applications, enhancing performance [33]. The careful selection and optimization of these hyper parameters can

lead to striking improvements in both accuracy and generalization, marking SVR as a formidable tool for regression tasks across a range of domains. Techniques such as cross-validation and grid search open the door to systematic exploration of this hyper parameter space, ensuring the identification of the most effective configuration—one that maximizes performance while guarding against overfitting [34].Incorporating this rigorous approach not only heightens the model’s predictive capabilities but also fosters a more profound comprehension of the data’s deeper patterns. This, in turn, empowers more informed decision-making, bringing tangible benefits in real-world scenarios.

ANALYSIS AND DISCUSSION

The data in Table 1 unveils a fascinating array of process parameters and their nuanced impact on biodiesel production yields. These parameters include free fatty acid (FFA) content, moisture percentage, viscosity, reaction time, and the corresponding biodiesel yield, each playing a pivotal role in the overall process.FFA content emerges as a particularly influential factor, with levels ranging from 1.7% to 3.5%. As one would expect, an increase in FFA often necessitates a more sophisticated transesterification process, but what is truly striking is the inverse relationship between FFA content and

biodiesel yield. At the lower end of the spectrum, the yield peaks at an impressive 89% when FFA is just 1.7%. However, as FFA content creeps up, especially towards the higher threshold of 3.5%, yield significantly drops to 78.5%. This suggests a clear trend: higher FFA levels may impair the efficiency of biodiesel production, pushing for more complex methods to achieve optimal results. Moisture content, while not as variable as FFA, spans from 0.05% to 0.3%. Despite this seemingly narrow range, moisture plays an intricate role. Though excess moisture can sabotage the transesterification reaction by disrupting catalyst activity, this dataset reveals little correlation between moisture fluctuations and yield changes. This might indicate that moisture, within these small variations, does not markedly hinder biodiesel production.Viscosity, another key variable, fluctuates between 35 and 43 cSt, yet its relationship with yield is far from straightforward. At first glance, higher viscosity values seem to accompany slightly higher yields, as seen with a viscosity of 42 cSt and a yield of 88%. Yet, this is not always the case—other instances of elevated viscosity, such as 41 cSt, coincide with more moderate yields. This suggests that viscosity, while certainly important, is likely not the dominant factor at play when other variables are accounted for.



**TABLE 1.**Biodiesel production process parameters and experimental yield %

FFA_(%)	Moisture_(%)	Viscosity_(cSt)	Time_(hrs)	Biofuel_Yield_(%)
2.5	0.1	40	2	85.5
3.2	0.2	35	2.5	80
1.8	0.05	42	3	88
2.1	0.15	38	2.2	83
3.5	0.3	37	2.8	79
2.3	0.12	41	3.2	87
2	0.08	39	2.4	84.5
2.7	0.18	36	2.6	81
1.9	0.1	43	3.1	89
3	0.25	37	2.5	82
2.6	0.08	41	2.7	82.5
1.7	0.12	38	3.3	88.5
2.2	0.14	39	2.3	84
3.1	0.19	36	2.4	80.5
2	0.09	42	3	87.5
2.4	0.11	40	2.5	83.5
3.3	0.21	37	2.6	79.5
1.9	0.1	43	3.1	89
2.5	0.13	38	2.2	84
3.4	0.24	36	2.7	78.5
2.8	0.17	39	3.2	86
2.1	0.1	42	2.1	85
3	0.22	35	2.5	80
1.7	0.09	41	3	88
2.2	0.12	38	2.2	83
3.3	0.19	37	2.8	79
1.8	0.08	42	3	88
2.3	0.11	40	2.4	84.5
3.4	0.24	37	2.7	78.5
2.8	0.17	39	3.2	86
2.2	0.1	42	2.1	85
3	0.22	35	2.5	80

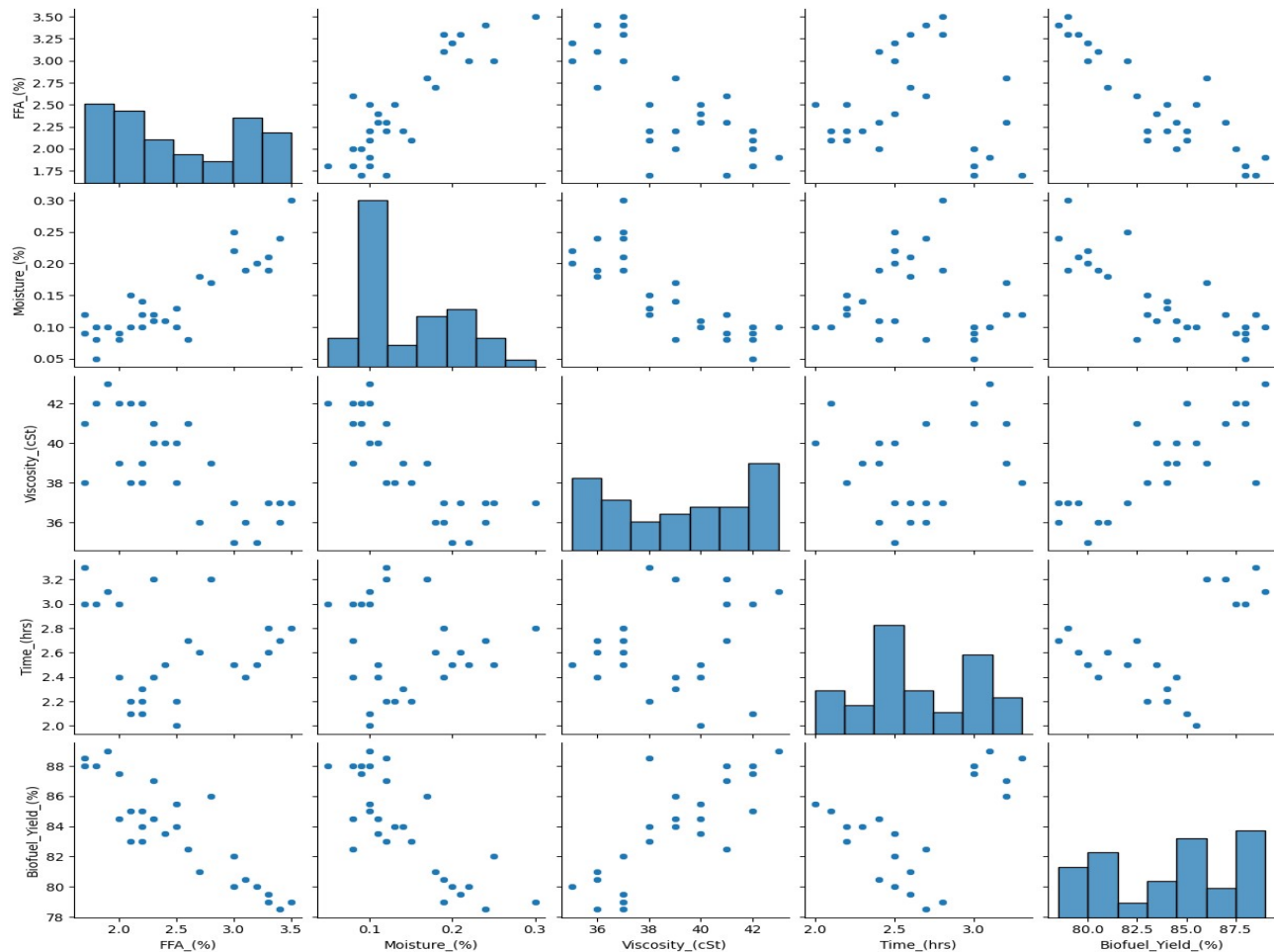
1.7	0.09	41	3	88
2.3	0.11	40	2.4	84.5
3.2	0.2	35	2.5	80
1.8	0.1	42	3	88
2.2	0.1	42	2.1	85
3	0.22	35	2.5	80
1.7	0.09	41	3	88
2.3	0.11	40	2.4	84.5
3.4	0.24	37	2.7	78.5
2.8	0.17	39	3.2	86
2.2	0.1	42	2.1	85
3	0.22	35	2.5	80
1.7	0.09	41	3	88
2.3	0.11	40	2.4	84.5
3.3	0.19	37	2.8	79
1.8	0.1	42	3	88
2.2	0.1	42	2.1	85
3	0.22	35	2.5	80

Finally, reaction time, which spans from 2 to 3.3 hours, emerges as a factor that warrants attention. The highest yield of 89% occurs at 3 hours, hinting at the possibility that longer reaction times, within this window, may foster better conversion to biodiesel. Overall, the data paints a picture of a complex,

dynamic interplay between process parameters, where higher FFA levels and extended reaction times tend to align with improved biodiesel yields, though other factors must not be overlooked in this intricate equation.

#### **Effect of Process Parameters:**





**FIGURE 1.** Satter plot of the various biofuel production process parameters and yield

The visualization presented in Figure 1—a correlation matrix—provides an intriguing glimpse into the intricate relationships between crucial parameters in the biofuel production process. Set against the backdrop of a 5x5 grid, this plot weaves together scatter plots and histograms, offering a dual perspective: one that showcases the distribution of individual parameters and another that delves into their pairwise interactions. The parameters explored here include Free Fatty Acid (FFA) content, moisture content, viscosity, reaction time, and, ultimately, the biofuel yield. On the diagonal, histograms of the individual parameters unfold, exposing their underlying distribution patterns.

Take, for instance, the FFA content: its distribution, marked by multiple peaks, hints at the existence of distinct groups or variability in process conditions across different batches or samples. The moisture content, on the other hand, follows a skewed distribution, where the bulk of values cluster towards the lower end, subtly indicating that lower moisture levels are

more prevalent within this dataset. Viscosity appears to follow a relatively uniform distribution, contrasting with the reaction time, which reveals more clustered behavior—suggesting certain time intervals dominate the process. As for the biofuel yield, its distribution is far from straightforward: multiple peaks dot the curve, hinting at a certain level of unpredictability or variability in production outcomes. The scatter plots off the diagonal dig deeper, offering a more nuanced perspective on the relationships between parameter pairs. A faint negative correlation between FFA content and biofuel yield becomes apparent, suggesting that lower levels of FFA may indeed favour higher yields. In contrast, the reaction time and yield display distinct clustering, pointing to specific time intervals that appear to optimise biodiesel production. However, the scatter between moisture content and yield remains scattered, resisting any clear linear trend, which may imply that moisture content doesn't have a direct influence on yield in the observed range.

Viscosity also shows a relationship with yield, but one that is anything but linear, hinting at more intricate and less obvious interactions at play. Taken together, this correlation matrix serves as a potent tool for dissecting parameter interactions and their collective impact on biofuel yield. The presence of non-linear patterns across many of the relationships highlights the

process’s inherent complexity, suggesting that more advanced models might be necessary for precise predictions. Ultimately, this visualization is invaluable, offering key insights into the parameters at play and laying the foundation for refining and optimizing biofuel production processes.

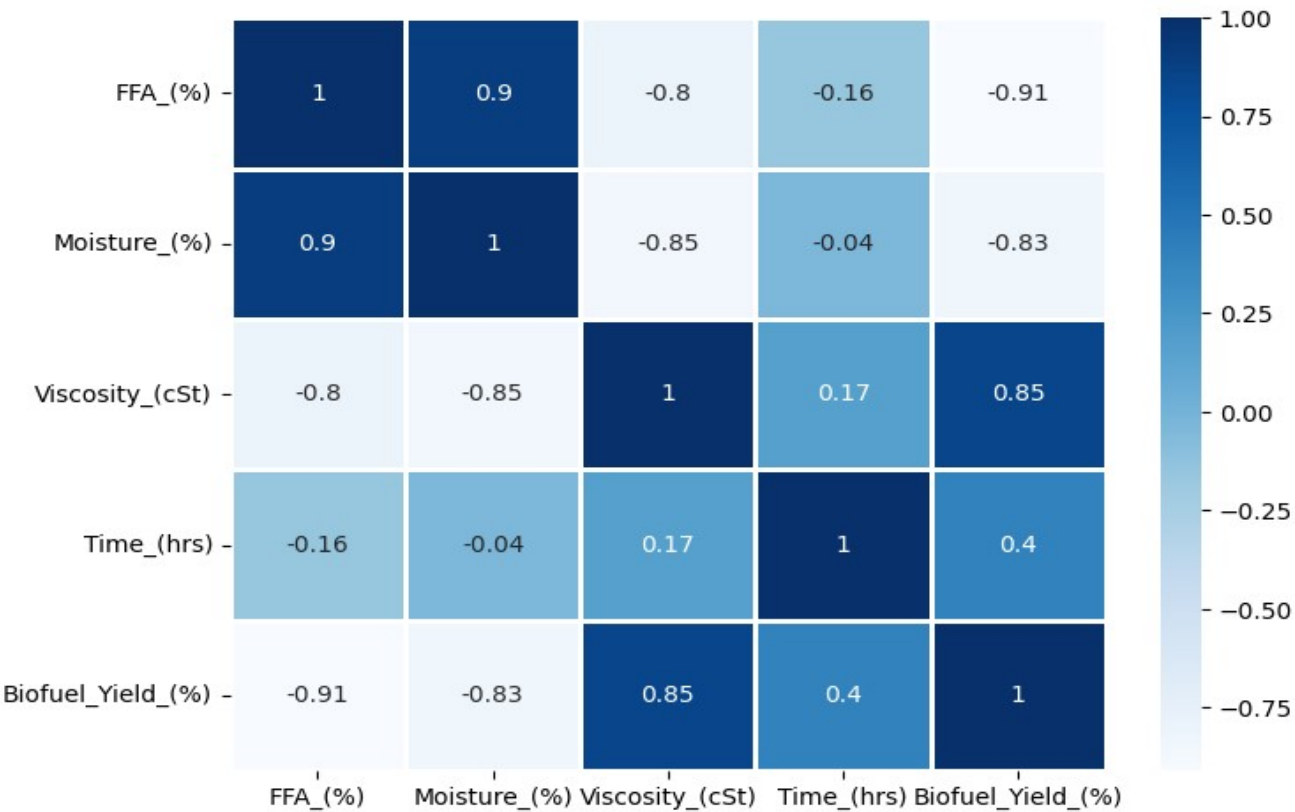


Figure 2. Correlation between the biofuel production process parameters and yield

The correlation matrix in figure 2 offers a visual representation of the intricate relationships between various parameters in the biofuel production process. It employs a colour-coded heat map, where the intensity of the blue shifts from dark to light, with darker hues indicating positive correlations and lighter tones reflecting negative ones. The correlation coefficients span from -1 to 1, with the colour intensity serving as a visual cue for the strength of these relationships. Among the most striking findings is the remarkable negative correlation (-0.91) between Free Fatty Acid (FFA) content and biofuel yield. As the FFA percentage rises, the biofuel yield takes a sharp dive, highlighting the detrimental impact of elevated FFA levels on the overall process. Similarly, moisture content is tightly bound to biofuel yield, with a strong negative correlation (-0.83), signifying that higher moisture

content compounds the problem, further reducing yield. In contrast, viscosity is positively correlated (0.85) with biofuel yield, suggesting that higher viscosity tends to coincide with better yields. Time, on the other hand, shows a more moderate positive correlation (0.4), indicating that a longer reaction period may slightly improve yield, but not to the same extent as the other parameters. Further inspection reveals noteworthy correlations among the process parameters themselves.

FFA and moisture content share an unexpectedly strong positive correlation (0.9), implying they often increase or decrease in tandem. Both FFA and moisture content exhibit substantial negative correlations with viscosity (-0.8 and -0.85, respectively), reinforcing the inverse relationship between these parameters. Reaction time, meanwhile, displays a rather weak connection to other factors, with correlation coefficients

fluctuating between -0.16 and 0.17, pointing to the relative independence of reaction time from the other process variables. These correlations reveal crucial insights for optimizing the production process. The robust negative correlations of FFA and moisture content with yield underscore the importance of minimizing these factors to boost productivity. The strong

positive link between viscosity and yield suggests that regulating viscosity could be pivotal in maximizing output. The moderate positive correlation with time indicates that, while longer reaction times may offer some benefits, they are not as critical as controlling FFA, moisture, or viscosity for optimal biofuel yields.

Linear Regression Model:

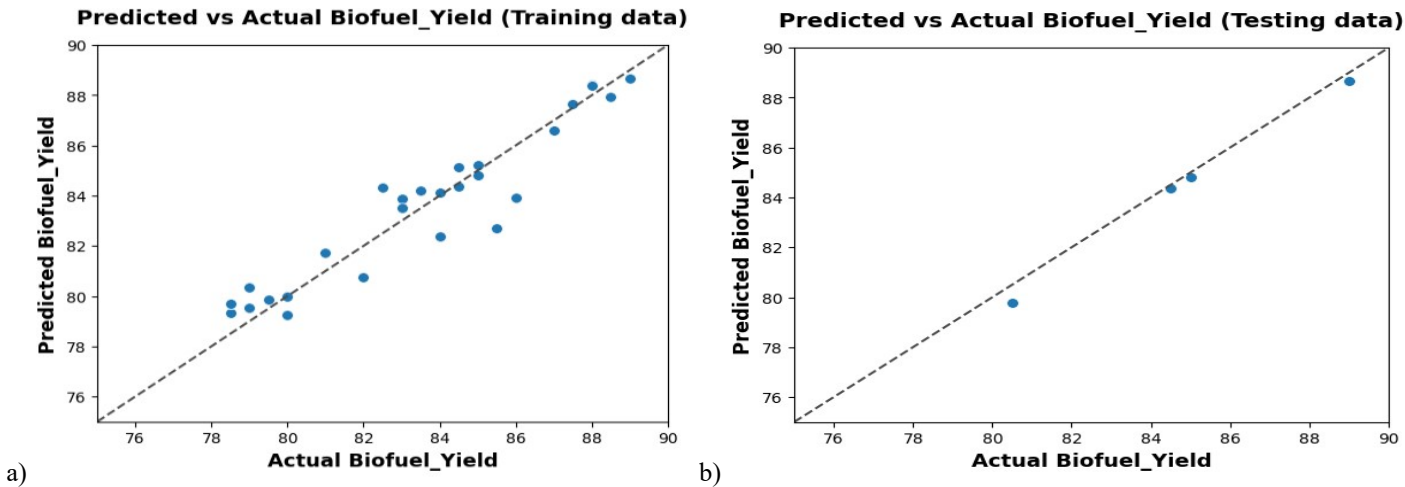


FIGURE 3.Linear Regression Model (training and testing)

Figures 3 a) and 3 b) present scatter plots that juxtapose predicted biodiesel yields against actual biodiesel yields, drawn from data supplied by a linear regression model. Figure 3 showcases the training data, whereas

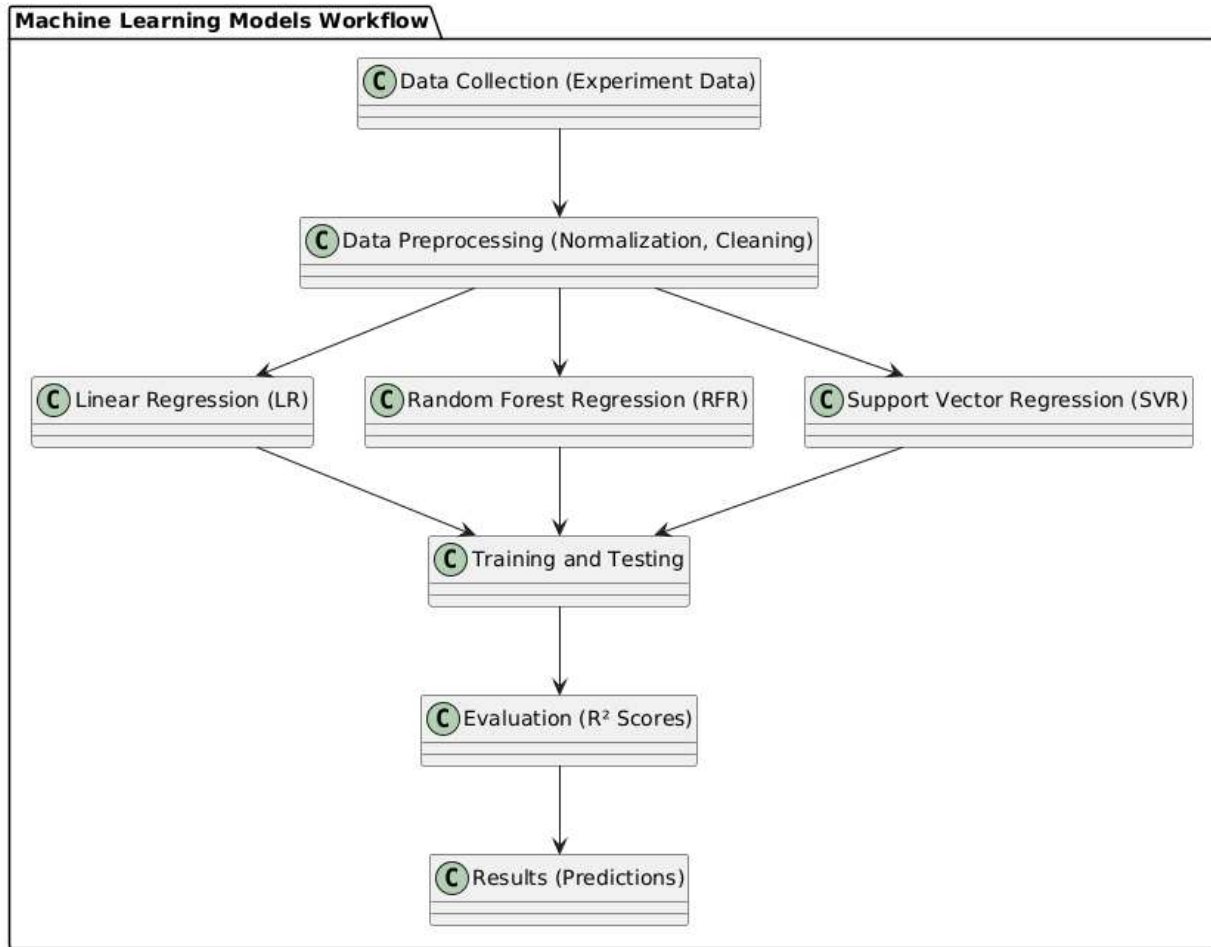


Figure 4 illustrates the testing data. The dashed line in both figures signifies the ideal scenario where predicted yields perfectly align with the actual yields. Each data point represents a specific observation, with its positioning reflecting the precision of the model's predictions. In Figure 3 a), which depicts the training data, the points generally hover near the dashed line. This proximity signals that the linear regression model has effectively captured the relationship between the input variables and biodiesel yield. The overall trend demonstrates a robust correlation between predicted and actual yields, with only a few points straying slightly. Such closeness to the ideal line suggests that the model has learned effectively from the training set, forecasting biodiesel yields with commendable accuracy. The compactness of the data points around the dashed line speaks to the model's success in minimizing errors during the training phase. Turning to Figure 3 b), the testing data mirrors this trend, with most points still clustered near the ideal dashed line. This suggests that the linear regression model retains its predictive prowess when confronted

with unseen data, generalizing well beyond the initial training set. However, there is a subtle increase in variation within the testing data compared to the training data. A handful of points in Figure 3 b) exhibit more pronounced deviations from the dashed line, implying that although the model is largely accurate, it's not infallible when predicting new data.

Such deviations are to be expected in practical scenarios, as even the most finely tuned models tend to show some level of error with fresh, unobserved data. The striking similarity between the two figures underscores the model's capacity to maintain its predictive strength from training to testing, a promising indicator of its robustness and ability to generalize. The model's aptitude for accurately predicting biodiesel yield based on parameters such as FFA, moisture, and viscosity suggests it holds significant potential for refining biodiesel production processes. Yet, the modest discrepancies observed in the testing data highlight an opportunity for refinement—whether through additional variables or further optimization—to enhance the models accuracy in future forecasts.

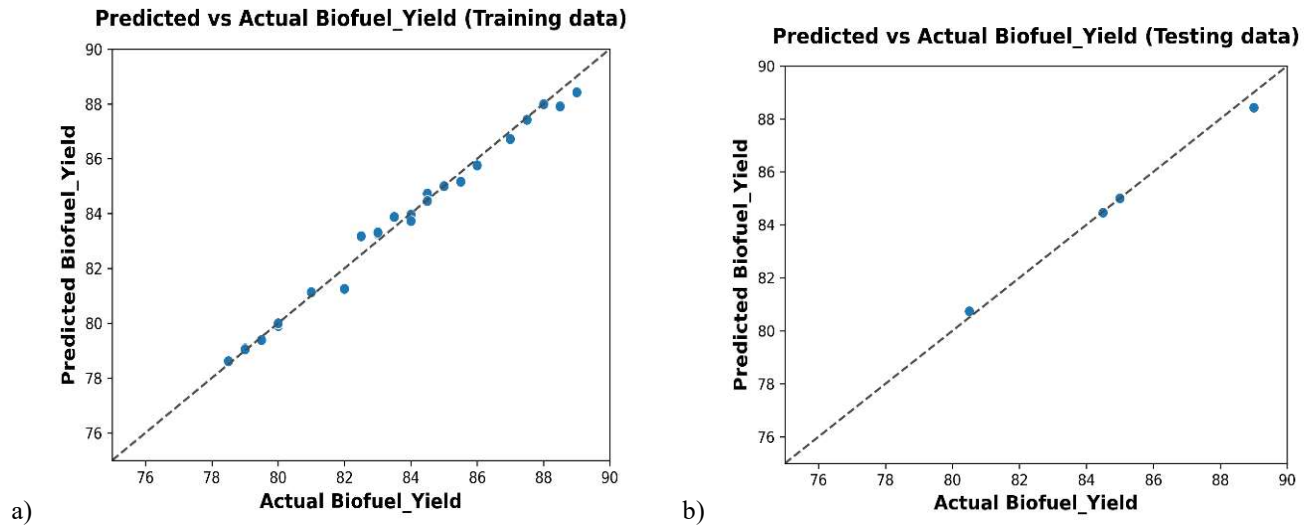
**Random Forest Regression:****FIGURE 4.**Random Forest Regression (training and testing)

Figure 4 presents the performance of a Random Forest Regression model in predicting biofuel yields, providing a comparison between predicted and actual values for both training and testing datasets. In the training data scatter plot (fig 4.a), a striking correlation emerges between the predicted and actual biofuel yields, with values spanning from approximately 76 to 90. The dashed diagonal line, symbolizing perfect predictions (where predicted equals actual), is closely followed by the blue dots—each representing individual predictions. This tight alignment signals that the Random Forest model has effectively grasped the patterns inherent in the training data. The clustering of points around the diagonal line is notably compact, suggesting that prediction error is minimal, with model accuracy peaking during the training phase. Interestingly, the uniform distribution of points across the entire range indicates a consistency in performance, regardless of yield values. When examining the testing data scatter plot (fig 4.b), we notice fewer points, which is typical of testing sets being smaller than their training counterparts. Nevertheless, the results remain equally

promising, with predictions continuing to hover near the diagonal line. Despite the smaller dataset, around 4-5 data points fall within the same range (80-90) as the training data. The closeness of these points to the diagonal is particularly noteworthy—it not only reinforces the model's adeptness at generalizing to new, unseen data but also suggests that the model is not overfitting to the training set. The Random Forest Regression model appears to excel in its biofuel yield prediction task, with its solid performance across both training and testing datasets. This consistency hints that the model is uncovering true data patterns rather than merely memorizing the training set. There are no significant outliers or systematic biases to speak of, as the points remain well-behaved, with no consistent deviation above or below the diagonal line. The model sustains its accuracy throughout the full range of yield values, demonstrating robustness across various input conditions. The parallel behavior between the training and testing plots strongly suggests excellent model generalization—an essential feature for applications in real-world biofuel production.



Support Vector Regression:

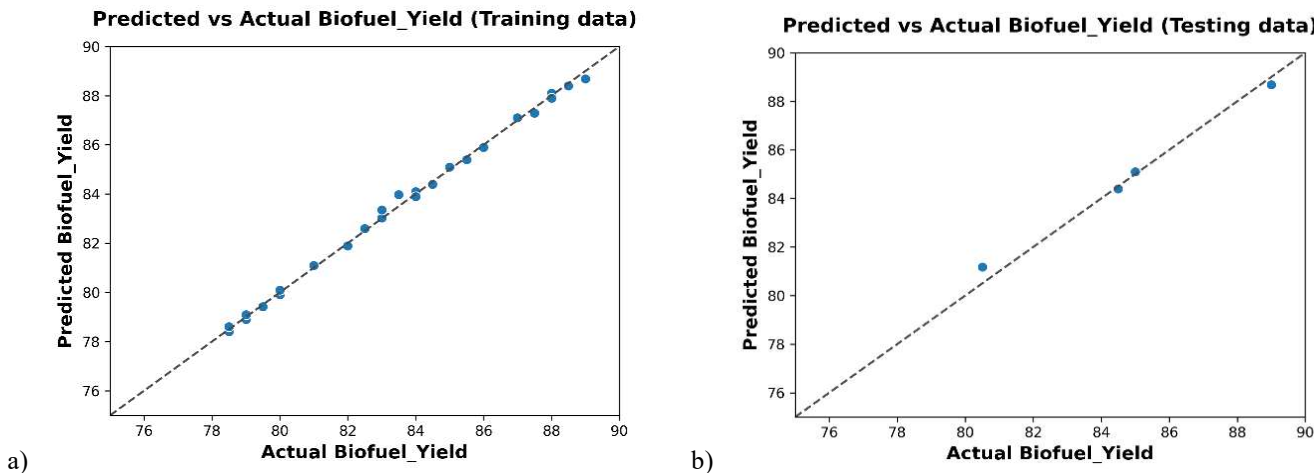


FIGURE 5.Support Vector Regression (training and testing)

Figure 5 presents the performance of a Support Vector Regression (SVR) model in predicting biofuel yields, offering distinct visual representations for both the training and testing datasets. Focusing on the training data plot (5.a), it becomes evident that the model excels in terms of predictive performance. The data points, shown in blue, are tightly clustered around the dashed diagonal line, which represents an ideal scenario where predicted values perfectly mirror actual values. With the biofuel yield spanning from roughly 76 to 90 units, the remarkable alignment of the points to the diagonal line over this entire range strongly suggests that the SVR model has adeptly captured the intricate relationships in the training data—without displaying significant bias at any specific yield level.

This indicates an impressive capacity to model the underlying patterns with precision. Turning to the testing data plot (5.b), which plays a crucial role in assessing how well the model generalises to unseen data, the picture is equally compelling. Fewer data points are present, as expected for a testing dataset, with only about three to four points visible, scattered across the range of 80 to 89 units. Yet, these points continue to adhere closely to the diagonal, suggesting that the model retains its predictive accuracy when confronted with new,

previously unseen data. This is particularly significant as it implies that the SVR model has avoided overfitting the training data, a common pitfall in machine learning. When compared to typical SVR performance, these results are highly commendable. SVR models are renowned for their ability to navigate non-linear relationships, maintaining robustness even in the presence of outliers—a quality evident here. The consistency of performance across both datasets points to the model's successful identification of the fundamental relationship between the input features and the biofuel yield.

However, a subtle limitation worth acknowledging lies in the relatively small number of testing points. This makes it more challenging to comprehensively validate the model's ability to generalise across the full spectrum of biofuel yields. Despite this, the available testing points align well with expected values, providing strong evidence that the model will perform reliably in practical scenarios. In conclusion, these visualisations reveal that the SVR model has successfully mastered biofuel yield prediction with notable accuracy, positioning it as a potentially invaluable tool for real-world biofuel production applications, where precise yield forecasting is essential for optimising production processes and guiding strategic planning.

TABLE 2.Performance Metrics of Regression Models

Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
------	--------	-------	----	-----	-----	------	-----	----------	------	-------

Train	LR	Linear Regression	0.920545	0.920545	0.928382	0.963526	0.699544	2.811881	0.000131	0.414163
Train	RFR	Random Forest Regression	0.995083	0.995159	0.057450	0.239688	0.146611	0.747500	0.000008	0.072500
Train	SVR	Support Vector Regression	0.998317	0.998329	0.019665	0.140232	0.118582	0.472697	0.000003	0.100023
Test	LR	Linear Regression	0.979986	0.992996	0.145304	0.381188	0.307337	0.729678	0.000021	0.176917
Test	RFR	Random Forest Regression	0.989171	0.990020	0.078619	0.280390	0.180500	0.572500	0.000010	0.037500
Test	SVR	Support Vector Regression	0.984005	0.984397	0.116124	0.340770	0.256578	0.675044	0.000017	0.099941

Table 2 offers a detailed comparison of three regression models—Linear Regression (LR), Random Forest Regression (RFR), and Support Vector Regression (SVR)—assessed across a range of performance metrics for both training and testing datasets, specifically in the context of biofuel yield prediction. On the training data, SVR stands out with the best overall performance, boasting an exceptional  $R^2$  of 0.998. Close behind, RFR achieves a solid 0.995, while LR lags at 0.920. The supremacy of SVR is evident, as it also records the lowest MSE (0.019), RMSE (0.140), and MSLE (0.000003), underscoring its precise predictions.

Moreover, the Max Error metric confirms SVR's reliability, showing the smallest deviation (0.472) compared to RFR's 0.747 and LR's significantly higher 2.811, further indicating SVR's robustness across varying data ranges. Shifting to the test data, all models exhibit strong predictive power, although their performance slightly diminishes relative to training. Notably, RFR emerges with the best generalization ability, reaching the

highest  $R^2$  of 0.989 on the test set, trailed by SVR (0.984) and LR (0.979). RFR also excels in error metrics, registering the lowest MSE (0.078), RMSE (0.280), and MAE (0.180), indicating its superior capacity to handle unseen data effectively. The Explained Variance Score (EVS) closely mirrors the  $R^2$  scores for all models, signalling that the models are accurately capturing the true patterns, rather than fitting noise.

Furthermore, the Median Absolute Error (MedAE) values remain consistently lower than the MAE for all models, pointing to the fact that while larger errors may occur, the typical deviations are relatively minor. In conclusion, while SVR delivers the best performance on training data, RFR demonstrates a remarkable ability to generalise, making it a more reliable choice for real-world applications. Meanwhile, Linear Regression, despite its simplicity, continues to perform commendably, suggesting that a strong linear relationship exists between the input features and biofuel yield.

## CONCLUSION

Biodiesel production from waste cooking oil (WCO) has emerged as a compelling and sustainable alternative to conventional biofuels. WCO, an often-overlooked by-product of the food industry, holds immense potential for conversion into biodiesel. This process not only addresses waste disposal concerns but also mitigates reliance on fossil fuels, presenting an environmentally friendly solution that contributes to the principles of a circular economy. A thorough examination of biodiesel production from WCO uncovers the remarkable role that machine learning techniques can play in optimizing this eco-conscious energy solution. By scrutinizing three distinct

regression models—Linear Regression (LR), Random Forest Regression (RFR), and Support Vector Regression (SVR)—we have unearthed profound insights into the predictive modeling of biodiesel yields, offering a clearer understanding of the intricate relationship between key process parameters and the outcomes of biodiesel production.

The comparative analysis of the model performance metrics uncovers an intriguing pattern: while all three models demonstrated robust predictive power, each revealed distinct strengths that set them apart. SVR, for instance, outshone the rest on the training data, achieving an impressive  $R^2$  value of 0.998, marking a near-perfect fit. Meanwhile, RFR excelled in

its ability to generalise to unseen data, securing a commendable  $R^2$  of 0.989 on the test set. Even the simpler Linear Regression model, often underestimated, made its mark with an  $R^2$  of 0.979 on the test data. This performance suggests a solid linear relationship between the process parameters and the biodiesel yield, albeit less intricate than its counterparts. Equally compelling was the study's insight into the pivotal role of process parameters in biodiesel production efficiency. The negative correlations between Free Fatty Acid (FFA) content, moisture levels, and biodiesel yield were striking, underscoring the critical need to carefully manage these parameters to avoid detrimental effects.

On the flip side, the positive relationship between viscosity and yield emerged as a key takeaway, hinting that regulating viscosity could be vital to optimizing production processes and enhancing yields. These findings carry profound implications for the large-scale production of biodiesel derived from WCO. The successful implementation of machine learning models in

predicting biodiesel yields presents a groundbreaking tool for process optimization and quality control. With the remarkable capacity of these models to forecast yields from specific input parameters, manufacturers can fine-tune their production processes, significantly reducing waste while simultaneously boosting efficiency. Looking ahead, this research unlocks multiple exciting prospects for further exploration. There is ample scope to integrate a broader array of process parameters, to delve deeper into alternative machine learning methodologies, and to develop real-time monitoring systems informed by these predictive models. Such advancements could dramatically elevate both the efficiency and dependability of biodiesel production from WCO. The successful deployment of these predictive tools holds the potential to play a pivotal role in the widespread adoption of WCO-based biodiesel, positioning it as a crucial, sustainable alternative to fossil fuels and advancing global initiatives for renewable energy and waste minimization.

## REFERENCES

1. Jackson, Robert B., Corinne Le Quéré, R. M. Andrew, Josep G. Canadell, Jan Ivar Korsbakken, Zhu Liu, Glen P. Peters, and Bo Zheng. "Global energy growth is outpacing decarbonization." *Environmental Research Letters* 13, no. 12 (2018): 120401.
2. Köpke, Michael, Steffi Noack, and Peter Dürre. *The past, present, and future of biofuels-Biobutanol as promising alternative*. Universität Ulm, 2011.
3. Gundekari, Sreedhar, Joyee Mitra, and Mohan Varkolu. "Classification, characterization, and properties of edible and non-edible biomass feedstocks." In *Advanced Functional Solid Catalysts for Biomass Valorization*, pp. 89-120. Elsevier, 2020.
4. Singh, Digambar, Dilip Sharma, S. L. Soni, Chandrapal Singh Inda, Sumit Sharma, Pushpendra Kumar Sharma, and Amit Jhalani. "A comprehensive review of biodiesel production from waste cooking oil and its use as fuel in compression ignition engines: 3rd generation cleaner feedstock." *Journal of Cleaner Production* 307 (2021): 127299.
5. Goh, Brandon Han Hoe, Cheng Tung Chong, Yuqi Ge, Hwai Chyuan Ong, Jo-Han Ng, Bo Tian, Veeramuthu Ashokkumar, Steven Lim, Tine Seljak, and Viktor Józsa. "Progress in utilisation of waste cooking oil for sustainable biodiesel and biojet fuel production." *Energy Conversion and Management* 223 (2020): 113296.
6. Sangkharak, Kanokphorn, Pimchanok Khaithongkaeo, Teeraphorn Chuaihunupakarn, Aopas Choonut, and Poonsuk Prasertsan. "The production of polyhydroxyalkanoate from waste cooking oil and its application in biofuel production." *Biomass Conversion and Biorefinery* 11 (2021): 1651-1664.
7. Gouran, Ashkan, Babak Aghel, and Farzad Nasirmanesh. "Biodiesel production from waste cooking oil using wheat bran ash as a sustainable biomass." *Fuel* 295 (2021): 120542.
8. Tsoutsos, T. D., S. Tournaki, O. Paraíba, and S. D. Kaminaris. "The Used Cooking Oil-to-biodiesel chain in Europe assessment of best practices and environmental performance." *Renewable and sustainable energy reviews* 54 (2016): 74-83.
9. Foteinis, Spyros, Efthalia Chatzisyneon, Alexandros Litinas, and Theocharis Tsoutsos. "Used-cooking-oil biodiesel: Life cycle assessment and comparison with first- and third-generation biofuel." *Renewable Energy* 153 (2020): 588-600.
10. Bozbas, Kahraman. "Biodiesel as an alternative motor fuel: Production and policies in the European Union." *Renewable and sustainable energy reviews* 12, no. 2 (2008): 542-552.
11. Roy, Tania, Shalini Sahani, Devarapaga Madhu, and Yogesh Chandra Sharma. "A clean approach of biodiesel production from waste cooking oil by using single phase BaSnO<sub>3</sub> as solid base catalyst: Mechanism, kinetics & E-study." *Journal of cleaner production* 265 (2020): 121440.
12. Sharma, Priyanka, Muhammad Usman, El-Sayed Salama, Margarita Redina, Nandini Thakur, and Xiangkai Li. "Evaluation of various waste cooking oils for biodiesel production: A comprehensive analysis of feedstock." *Waste Management* 136 (2021): 219-229.
13. Liu, Yanbing, Xinglin Yang, Abdullahi Adamu, and Zongyuan Zhu. "Economic evaluation and production process simulation of biodiesel production from waste

- cooking oil." *Current Research in Green and Sustainable Chemistry* 4 (2021): 100091.
14. Pugazhendhi, Arivalagan, Avinash Alagumalai, Thangavel Mathimani, and A. E. Atabani. "Optimization, kinetic and thermodynamic studies on sustainable biodiesel production from waste cooking oil: An Indian perspective." *Fuel* 273 (2020): 117725.
15. Basumatary, Sanjay, Biswajit Nath, and Pranjal Kalita. "Application of agro-waste derived materials as heterogeneous base catalysts for biodiesel synthesis." *Journal of Renewable and Sustainable Energy* 10, no. 4 (2018).
16. Chung, Zheng Lit, Yie Hua Tan, Yen San Chan, Jibrail Kansedo, N. M. Mubarak, Mostafa Ghasemi, and Mohammad Omar Abdullah. "Life cycle assessment of waste cooking oil for biodiesel production using waste chicken eggshell derived CaO as catalyst via transesterification." *Biocatalysis and Agricultural Biotechnology* 21 (2019): 101317.
17. Mohamed, R. M., G. A. Kadry, H. A. Abdel-Samad, and M. E. Awad. "High operative heterogeneous catalyst in biodiesel production from waste cooking oil." *Egyptian Journal of Petroleum* 29, no. 1 (2020): 59-65.
18. Su, Xiaogang, Xin Yan, and Chih-Ling Tsai. "Linear regression." *Wiley Interdisciplinary Reviews: Computational Statistics* 4, no. 3 (2012): 275-294.
19. Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.
20. James, Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. "Linear regression." In *An introduction to statistical learning: With applications in python*, pp. 69-134. Cham: Springer International Publishing, 2023.
21. Poole, Michael A., and Patrick N. O'Farrell. "The assumptions of the linear regression model." *Transactions of the Institute of British Geographers* (1971): 145-158.
22. Maulud, Dastan, and Adnan M. Abdulazeez. "A review on linear regression comprehensive in machine learning." *Journal of Applied Science and Technology Trends* 1, no. 2 (2020): 140-147.
23. Svetnik, Vladimir, Andy Liaw, Christopher Tong, J. Christopher Culberson, Robert P. Sheridan, and Bradley P. Feuston. "Random forest: a classification and regression tool for compound classification and QSAR modeling." *Journal of chemical information and computer sciences* 43, no. 6 (2003): 1947-1958.
24. Moorthy, Kohbalan, and Mohd Saberi Mohamad. *Random forest for gene selection and microarray data classification*. Springer Berlin Heidelberg, 2012.
25. Hengl, Tomislav, Madlene Nussbaum, Marvin N. Wright, Gerard BM Heuvelink, and Benedikt Gräler. "Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables." *PeerJ* 6 (2018): e5518.
26. Lingjun, He, Richard A. Levine, Juanjuan Fan, Joshua Beemer, and Jeanne Stronach. "Random forest as a predictive analytics alternative to regression in institutional research." *Practical Assessment, Research, and Evaluation* 23, no. 1 (2019): 1.
27. Dong, Penghao, Huachen Peng, Xianqiang Cheng, Yan Xing, Xin Zhou, and Dongliang Huang. "A random forest regression model for predicting residual stresses and cutting forces introduced by turning in 718 alloy." In *2019 IEEE International Conference on Computation, Communication and Engineering (ICCCE)*, pp. 5-8. IEEE, 2019.
28. Popuri, Sai K. "An Approximation Method for Fitted Random Forests." *arXiv preprint arXiv:2207.02184* (2022).
29. Lin, Pao-Tsun, Shun-Feng Su, and Tsu-Tian Lee. "Support vector regression performance analysis and systematic parameter selection." In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 2, pp. 877-882. IEEE, 2005.
30. Tong, Hongzhi, Di-Rong Chen, and Lizhong Peng. "Analysis of support vector machines regression." *Foundations of Computational Mathematics* 9, no. 2 (2009): 243-257.
31. Miao, Yu, Hongye Su, Ouguan Xu, and Jian Chu. "Support vector regression approach for simultaneous data reconciliation and gross error or outlier detection." *Industrial & engineering chemistry research* 48, no. 24 (2009): 10903-10911.
32. Takeda, Akiko, Jun-ya Gotoh, and Masashi Sugiyama. "Support vector regression as conditional value-at-risk minimization with application to financial time-series analysis." In *2010 IEEE International Workshop on Machine Learning for Signal Processing*, pp. 118-123. IEEE, 2010.
33. SUN, Deshan, Jinpei WU, Zhenting HOU, and Jianhua XIAO. "Support vector regression algorithm based on single parameter." *Journal of system engineering* 20, no. 1 (2005): 109-112.
34. Fisher, Dustin M., Ralph F. Kelly, Deep R. Patel, and Mark Gilmore. "A support vector regression method for efficiently determining neutral profiles from laser induced fluorescence data." *Review of Scientific Instruments* 89, no. 10 (2018).
35. Pendyala, S. K. *Transformation of Healthcare Analytics: Cloud-Powered Solutions with Data Science, ML, and LLMs*. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, 10(6), 724-734. Available at: <https://ijsrcseit.com/index.php/home/article/view/CSEIT241061114>
36. Pendyala, S. K. (2024). *Enhancing Healthcare Pricing Transparency: A Machine Learning and AI-Driven Approach to Pricing Strategies and Analytics*. *International*

- Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), 10(6), 2334-2344. Available at: <https://ijsrcseit.com/index.php/home/article/view/CSEIT2410612436>
37. Pendyala, S. K. (2024). Optimizing Cloud Solutions: Streamlining Healthcare Data Lakes For Cost Efficiency. *Technology (IJRCAIT)*, 7(2). Available at: [https://iaeme.com/MasterAdmin/Journal\\_uploads/IJRCAIT/VOLUME\\_7\\_ISSUE\\_2/IJRCAIT\\_07\\_02\\_113.pdf](https://iaeme.com/MasterAdmin/Journal_uploads/IJRCAIT/VOLUME_7_ISSUE_2/IJRCAIT_07_02_113.pdf)
38. Pendyala, S. K. . (2025). Data Engineering At Scale: Streaming Analytics With Cloud And Apache Spark. *Journal of Artificial Intelligence and Machine Learning*, 3(1), 1-9. <https://doi.org/10.55124/jaim.v3i1.248>
39. Adari, Vijay Kumar, "Interoperability and Data Modernization: Building a Connected Banking Ecosystem," *International Journal of Computer Engineering and Technology (IJCET)*, vol. 15, no. 6, pp. 653-662, Nov-Dec 2024. DOI: <https://doi.org/10.5281/zenodo.14219429>.
40. Adari, Vijay Kumar., Vinay Kumar, Ch., Srinivas, G., Kishor Kumar, A., & Praveen Kumar, K. (2024). Artificial neural network in fibre-reinforced polymer composites using the ARAS method. *SOJ Materials Science and Engineering*, 10(1), 1-11
41. Adari, Vijay Kumar., Vinay Kumar, Ch., Srinivas, G., Kishor Kumar, A., & Praveen Kumar, K. (2020). Explainability and interpretability in machine learning models. *Journal of Computer Science and Applied Information Technology*, 5(1), 1-7. <https://doi.org/10.15226/2474-9257/5/1/00148>
42. Praveen Kumar, K., Adari, Vijay Kumar., Vinay Kumar, Ch., Srinivas, G., & Kishor Kumar, A. (2024). Optimizing network function virtualization: A comprehensive performance analysis of hardware-accelerated solutions. *SOJ Materials Science and Engineering*, 10(1), 1-10.