Applied Chemical Engineering (2025) Volume 8 Issue 4 ISSN: 2578-2010 (O)
doi: 10.59429/ace.v8i4.5841

ORIGINAL RESEARCH ARTICLE

Data-Driven Prediction of Biofuel Yield and Combustion Emissions Using AI Techniques

Pallavi Vishnu Kharat[1], Beena Nawghare[2,9], N. Alangudi Balaji[3], Vishvas V. Kalunge[4], Charu P. Kumbhare[5], Tejasvini Rahul Katkar[6], Sagar Arjun Dalvi[7], Shital Yashwant Waware[8,9], Anant Sidhappa Kurhade[8,9] *

[1] *Civil Engineering Department, Ajeenkya D. Y. Patil School of Engineering, Lohegaon, Charoli Budruk, Pune - 412105, Maharashtra, India.*

[2] *Department of First Year Engineering (Engineering Chemistry), Dr. D. Y. Patil Institute of Technology, Sant Tukaram Nagar, Pimpri, Pune, 411018, Maharashtra, India*

[3] *Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur – 522502, Andhra Pradesh, , India*

[4] *Department of Computer Engineering, Dhole Patil College of Engineering, Wagholi, Pune – 412207, Maharashtra, India.*

[5] *Department of Mathematics, Ramdeobaba University, Ramdeo Tekdi, Katol Road, Nagpur - 440013, Maharashtra, India*

[6] *Department of Electronics and Telecommunication Engineering, Genba Sopanrao Moze College of Engineering, Balewadi, Pune - 411045, Maharashtra, India.*

[7] *Department of Mechanical Engineering, Dr. Bapuji Salunkhe Institute of Engineering & Technology (BSIET), Tarabai Park, Kolhapur 416003, Maharashtra, India*

[8] *Department of Mechanical Engineering, Dr. D. Y. Patil Institute of Technology, Sant Tukaram Nagar, Pimpri - 411018, Pune, Maharashtra, India*

[9] *Dnyaan Prasad Global University (DPGU), School of Technology and Research - Dr. D. Y. Patil Unitech Society, Sant Tukaram Nagar, Pimpri - 411018, Pune, Maharashtra, India*

***Corresponding author:** Anant Sidhappa Kurhade; a.kurhade@gmail.com

ARTICLE INFO

Received: 25 November 2025
Accepted: 16 December 2025
Available online: 25 December 2025

ABSTRACT

Accurate prediction of biofuel yield and combustion emissions plays a key role in improving conversion efficiency and reducing dependence on trial-and-error experiments. Biofuel systems involve diverse biomass feedstocks and complex thermochemical and combustion processes, which makes modeling difficult. Reliable prediction tools also support cleaner energy practices and informed process control. Existing research shows several clear limitations. Many studies rely on small, single-site datasets, which limit broader applicability. Data preprocessing methods differ widely across publications, leading to inconsistencies in reported results. Validation strategies are often limited to internal testing, which restricts confidence in real-world use. These issues reduce model generalization, reproducibility, and clarity of interpretation. This review examines recent progress in artificial intelligence and machine learning applied to biofuel production and engine emission prediction. It summarizes commonly used data sources, including laboratory experiments and engine tests. The review outlines feature selection and transformation methods adopted in prior work. It also reviews model construction strategies and evaluation practices used to assess performance. Surveyed studies show that ensemble learning methods, neural networks, and physics-informed hybrid models achieve

high prediction accuracy at laboratory scale. These models perform well for yield and emission estimation under controlled conditions. At the same time, several persistent challenges remain. Many advanced models show weak extrapolation beyond training ranges. Model transparency is also limited, which affects trust and interpretability. The findings indicate that benchmark datasets and consistent preprocessing protocols are needed. Strong external validation can improve reliability. Incorporating physical constraints into machine learning workflows can enhance stability and realism. Such practices can support real-time implementation and promote wider use of data-driven prediction tools in biofuel research and industrial operations.
Keywords: biofuel yield; combustion emissions; machine learning; deep learning; hybrid models; physics-informed ML; soft sensors

1. Introduction

The precise prediction of biofuel yield, as well as combustion-related emissions, is critical for process optimization, LCA (life-cycle assessment), and regulatory compliances. The traditional first-principles-based models still have their virtues, but they usually have difficulty in capturing the complexity and diversity of biological feedstocks as well as multiscale combustion phenomena. Data-driven AI methods enhance such models by directly learning nonlinear relationships from experimental and in-service data48. These approaches are useful for soft sensing, optimization and real process control. Recent studies show an explosive growth in the application of artificial neural networks, random forests, support vector machines, boosting techniques and deep learning approaches. These models are used to estimate biofuel yield and emission properties at lab scale and pilot scale with good predictive accuracy.

Increasing global energy demands and environmental concerns on the use of fossil fuels have lead to growing attention in renewable energy sources [1]. Biofuels are attracting significant attention as potential substitutes to mitigate greenhouse gas emissions and enhance energy security [2]. However, wide diversity in biomass composition as well as complexity of conversion and combustion processes add to the complication for cost effective production and reducing emissions [3]. More advanced predictive models are needed. AI and machine learning applications are widely used in the bioenergy chain to aid prediction, optimization and control [4]. In this review, we discuss the approach being taken to estimate biofuel yield and emissions prediction using AI/ML technologies. It presents practicability, technological challenges and the direction of future research as well. Correspondingly, the need for biomass-derived energy is increasing as a result of concerns about increasing greenhouse gas production [1]. AI driven models have been successful in addressing feedstock variability, cost of conversion reduction and enhance the reliability of supply chain [4–5]. Such models are accurate in representing complex processes such as pyrolysis, gasification and combustion [6]. Various machine learning approaches have been applied to predict bio-oil and gas yield in response to different feedstock and operational conditions [7,8]. These methods optimize the parameters, lower the synthesis cost and reduce the waste production [9]. They have the advantage over classical kinetic models that they can take into account nonlinear input–output relations [8]. Incorporation of multiple data sources also enhances predictive power in bioenergy systems [4].

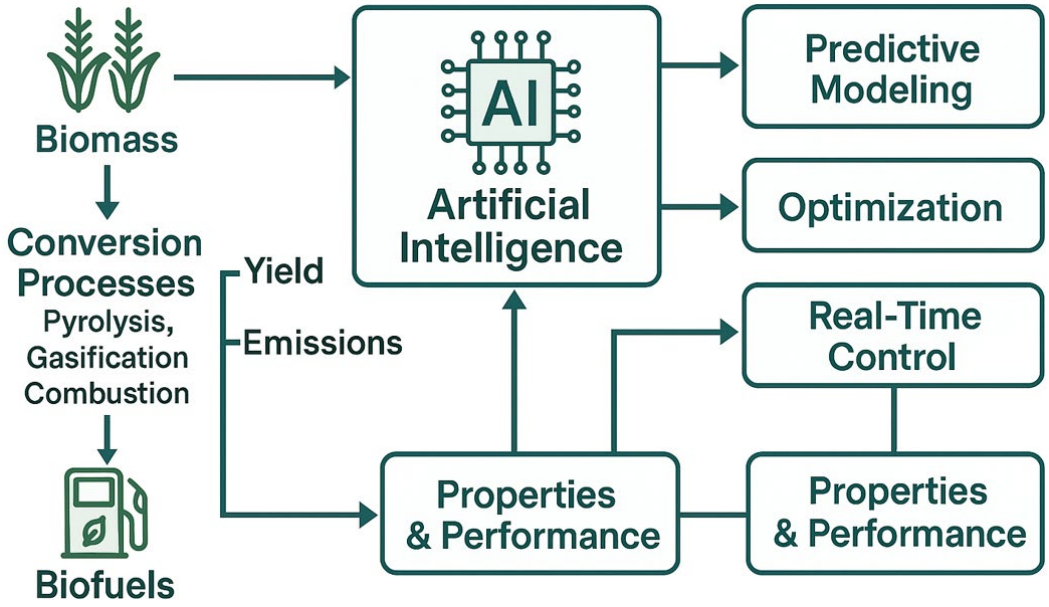

Figure 1. AI framework for predicting and improving biofuel yield and emissions.

Figure 1 shows conceptual relationship of AI and biomass conversion pathways for biofuel yield and emissions prediction. It emphasizes the role of AI in modeling, design, control and testing of biofuel characteristics. Machine learning can help in predicting fuel properties as well as end-use performance, leading to better design and operation of bioenergy systems [4]. The Ensemble Learning methods like Random Forest work well to predict bio-oil yield from pyrolysis with high determination coefficients [6]. These models also predict combustion products, a crucial issue for complying with stringent environmental regulations and obtaining cleaner fuels [10]. Accurate prediction also reduces the number of experiments that must be conducted, with time and resource savings in research and development activities [11]. AI-PM enhances biomass combustion system reliability and stability, extending its operating cycle [12]. Machine learning makes it possible to gain insight into complex biological processes through the exploration of pathways with optimal processing, minimization of energy consumption, and minimization of waste [2]. ML techniques have also been used to screen for appropriate gasification pathways based on limited data [13]. Biomass resource management and supply chain planning is another domain where AI finds application [2]. These input–output relation predictions are laid out by a data-driven trend analysis, forecasting how the input-explanatory parameters affect outputs including biochar yield and carbon content [14]. This is beneficial on the efficiency of biomass conversion and production of value-added biochar for soil amendment, carbon capture [6,14]. Models, such as XGBoost, for biochar yield and carbon content are good estimate to aid feedstock and operational decisions [14,15]. AI realizes full-chain optimization to minimize waste, from feedstock to conversion process [2]. strain selection, techno-economic analysis and life-cycle assessment is also supported [12]. Table 1 explains the key themes.

In conclusion, machine learning leads to rapid process optimization, minimizes experimental burden and speeds up next-generation biofuel development at laboratory and industrial scales.

Table 1. Summary of Key Themes in the Introduction

Aspect	Core Focus	AI/ML Techniques Mentioned	Bioenergy Processes Covered	Key Advantages Reported	Notable Insights from Literature
Need for Accurate Prediction	Importance of predicting biofuel yield and emissions	ANN, Random Forest, SVM, Gradient Boosting, Deep Neural Networks	Combustion, pyrolysis, gasification	Better handling of nonlinear relations; improved control	Rapid rise in AI studies at lab and pilot scales
Challenges in Biofuel Systems	Variability in feedstock and complex conversion	Hybrid ML–physics models, ensemble methods	Biomass conversion, reactor operations	Improved extrapolation; fewer assumptions	Key barriers include feedstock variability
AI for Yield Prediction	Modeling yields of bio-oil, biochar, and gases	Random Forest, XGBoost, SVR, ANN	Pyrolysis, gasification, fermentation	High R^2; optimal parameter identification	Many models achieve $R^2 > 0.85$
AI for Emission Forecasting	Predicting NOx, CO, CO_2, particulates	Gradient Boosting, Deep Learning	Engine and reactor combustion	Supports compliance; fewer experiments	Cuts time and cost of combustion testing
Process and Supply Chain Optimization	Efficiency in sourcing, conversion, logistics	Active learning, meta-learning, semi-supervised learning	Crop management, logistics, transesterification	Better resource use, reduced cost	Yield gains of 10–20% in ML-guided farming
Advanced Applications in Bioenergy	Predicting pathways, metabolic fluxes, fuel quality	ANN–RSM hybrids, ML-based experimental design	Microalgae, biodiesel systems	Higher yields (84–98%); fewer experiments	Validated ML accuracy for reaction outputs

2. Data types, features and pre-processing

2.1. Typical data sources

Research on TPD-based prediction of biofuel yield and combustion emission is based on data from various steps in biomass fuel production and use. Experiments are mainly based on laboratory-scale data. Such investigations provide the details of reaction conditions viz., temperature; alcohol/oil ratio, catalyst quantity, stirring speed, and reaction time for transesterification. In pyrolysis investigations, the most important variables are heating rate, reactor temperature, particle size of biomass, and type of catalyst. There are a number of feedstock-specific properties like moisture content, free fatty acid level, elemental composition as well as ash content which have been taken into the TAB by model or mapping as they significantly impact the conversion characteristics. Data provided by the engine test benches are necessary for the combustion and emission analysis. These data contain information on fuel consumption, engine efficiency, and exhaust emissions (specifically NOx, CO and unburned hydrocarbons) at various loads and speeds [21]. The experience from biorefineries and industrial agriculture is informative about the issues at scale-up, including feedstock variation, environmental impact, and equipment performance [27]. Pilot-plant and semi-industrial installations deliver dynamic data streams for sustained operation. Temperature, pressure, mass flow rate, residence time and intermediate compositions are measured by sensors. This kind of data may be considered better representative of actual operating conditions and transients not observed in laboratory tests [5]. Satellite and drone-based remote sensing activities have been advocated for gross feedstock evaluation, crop monitoring, and the estimation of biomass yields on large scale [31]. Public databases and literature report average biomass composition, conversion yield and environmental performance [32]. Engine testing also comprises stationary and transient cycles, through the use of inputs such as fuel properties and operational conditions. CFD simulations contribute virtual data on in-cylinder phenomena and combustion kinematics [33]. There are issues

related to the measurement uncertainty that need to be resolved in order for interpretation to be reliable [34]. Derived physicochemical properties describe the relationship between feedstock quality and performance. Model transfer is to a great extent determined by the quality and diversity of data [35–37]. The literature review indicates that the preferred approach for model calibration and validation is highly influenced by laboratory studies and engine/pilot scale measurements. However, remote sensing–derived information such as biomass availability (from satellite or drone data), crop health indices or spatial variabilities are currently only sparsely used and mainly limited to feedstock assessment rather than yield / emission prediction. Long-term datasets in this context of industrial biorefineries or even continuous plants on several sites are also rare. This imbalance inhibits model generalizability over different regions, seasons or supply chain systems and hinders the seamless embedding of AI models in real-world system-level decision-making.

2.2. Feature engineering and preprocessing

Data pre-processing is an iterative activity that involves data cleaning and feature engineering. Numerical variables are commonly normalized or standardized to reduce scale-related bias during model training. Categorical inputs such as feedstock type, catalyst, reactor configuration, and operating mode are converted into numerical form using techniques like label encoding or one-hot encoding. Missing values are treated through imputation methods, including mean, median, mode, or advanced machine learning–based approaches, to maintain dataset consistency. Outlier detection and treatment are also necessary since extreme values can strongly influence model performance. Feature selection and dimensionality reduction techniques such as Principal Component Analysis, recursive feature elimination, and permutation importance are applied to identify influential variables while reducing feature space, which improves interpretability and reduces computational cost [1,18]. Figure 2 explains the Typical Data Sources for Biofuel Yield and Emission Modelling.

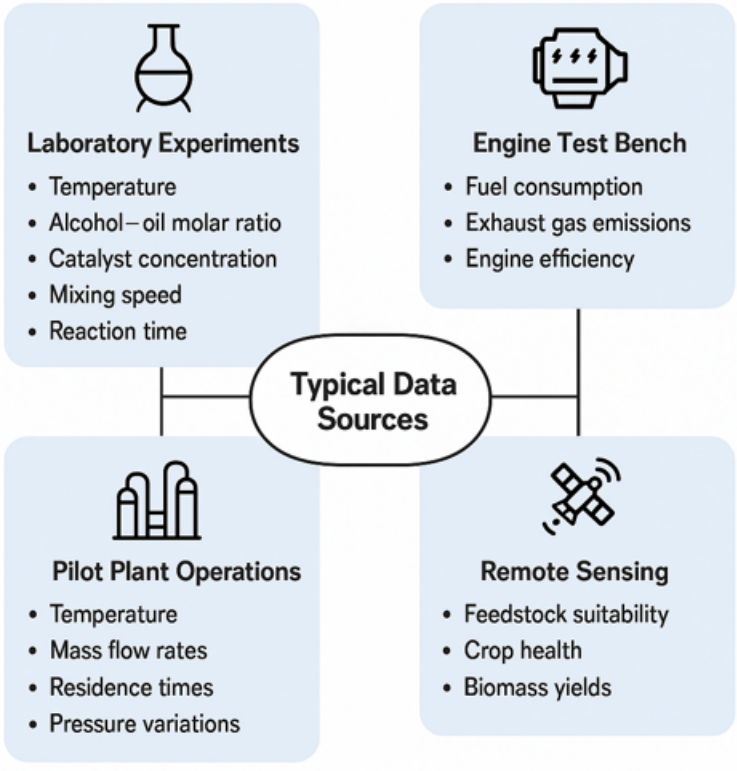

Figure 2. Typical Data Sources for Biofuel Yield and Emission Modelling

In biofuel research, data scarcity is a common limitation, leading to the use of algorithms suitable for small datasets or the integration of synthetic and experimental data through multi-fidelity modeling strategies [12]. When datasets contain many correlated physicochemical variables, dimensionality reduction methods like

PCA are preferred to retain essential information while simplifying the feature space. Mechanistic model outputs or pseudo-samples are sometimes added to enhance model robustness. Proper handling of variability across feedstocks, reactor types, and operating conditions supports the transferability of machine learning models to stable biofuel production scenarios [1,38]. Missing data often arise from measurement limitations or equipment issues, and are addressed using mean or median substitution, regression-based imputation, or k-nearest neighbour methods. Reliable model evaluation requires appropriate validation strategies, particularly for small datasets, where k-fold or nested k-fold cross-validation is recommended. Incorrect selection of performance metrics can lead to misleading conclusions regarding model generalization [39]. Overall, careful data preprocessing, validation, and dataset design are essential for developing robust and transferable AI models in biofuel applications [18,40].

To improve clarity, the feature engineering and preprocessing steps can be viewed as a sequential workflow. First, raw datasets are screened for completeness, and missing values are addressed using simple statistical imputation (mean, median, or mode) or learning-based methods when variable interactions are important. Second, outlier detection and treatment are performed to reduce the influence of extreme or erroneous measurements, using threshold-based rules or distribution-aware methods. Third, categorical variables such as feedstock type, catalyst, and reactor configuration are encoded into numerical form through label encoding or one-hot encoding, depending on model requirements. Finally, numerical features are scaled using normalization or standardization to ensure balanced learning, followed by feature selection or dimensionality reduction to retain the most influential variables. This ordered preprocessing sequence clarifies data preparation and supports consistent and reproducible model development.

3. Machine learning approaches—overview and comparative strengths

3.1. Classical and ensemble methods

Classical machine-learning techniques are still prominently used for yield prediction and combustion emissions studies, as it performs well on structured tabular data which is often synonymous with this domain. Linear regression and partial least squares are commonly used to express linear relationships between process inputs and outputs, since these methods provide easy interpretability for the user that can assist in identifying key reaction or combustion parameters. They can manage the linear and nonlinear patterns, and they generalize well even with low data by building optimal decision boundaries or regression functions [8]. Decision trees and random forests offer transparent rules (easy to interpret) that serve well for mixed type data and complex non-linear relationships [18]. Ensemble algorithms such as Random Forest can create more robust and precise models by integrating multiple decision trees, resulting in lower overfitting and better prediction generalization [41,42]. Kernelized SVMs additionally expand the modeling power through non-linear transformations and often work well on medium-sized data sets. Tree-based and ensemble algorithms have lower preprocessing needs and can also innately embody feature interactions. Gradient boosting approaches such as XGBoost and LightGBM iteratively update predictions by learning from residual errors in most crop yield and emission estimation problems. CatBoost, its strength is categorial-variable-accelerated and reduces overfit with ordered boosting strategy [21]. In general, when there are uncertainties like noise, multicollinearity and complex feedstock–process interactions, the ensemble models via random forests (RF) and boosting techniques feature strong robustness and they obviously outperform regression-based predictions in biodiesel conversion, pyrolysis products prediction and engine emissions analysis.

3.2. Artificial neural networks and deep learning

ANNs are commonly employed when input—output relationships are highly nonlinear. A feedforward model and a multilayer perceptron are frequently used for the prediction of ester conversion, for optimizing the reaction conditions, or even in estimation of combustion emissions. These models approximate complex

relationships between variables and perform satisfactorily when large datasets from experiments or operations can be exploited. More recently, research has approached deeper and deeper architectures where convolutional neural networks process spectral data or chromatographic signals (and more generally structured inputs) as long short-term memory networks are now appropriate to time-dependent datasets such as for transient engine cycles. Deep models possess great learning power, but their large complexity and the overfitting effects often require either big data or ancillary approaches like transfer learning, synthetic data generation or physics-guided pretraining to avoid this degradation. Hybrid and ensemble models that integrate artificial neural networks with techniques such as Random Forest, XGBoost have demonstrated enhanced predictive performance and better generalization power, especially for challenging cases of urban vehicular emissions [43]. Meanwhile, a number of studies indicate that they are performing worse than simpler models (such as Elastic-Net and XGBoost) in some tasks like for NOx prediction [44]. Because of the nonlinear character, artificial neural networks (ANNs) are appropriate tools for such complex thermochemical processes as biomass gasification and pyrolysis [5]. The knowledge of local and global patterns they can learn qualify them as an effective tool for large biodiesel dataset with numerous input variables [18]. Multi-input multi-output deep (Nelson et al.) networks as well find application since they are able to predict on the fly simultaneously biofuel yield and quality indicators such as viscosity and heating value of the produced liquid, thus avoiding having multiple separate models and capturing dependencies among targets.

3.3. Hybrid and physics-informed machine learning (PIML)

Hybrid methods combine machine learning with mechanistic or semi-empirical models to enhance the ability of prediction for biofuel systems. In most industries mechanistic production models are existing where the model parameters are hard to determine directly. It is employed to estimate missing parameters, learn residual corrections or infer unobserved states, the latter sometimes in the context of a soft sensor. Research on biodiesel transesterification and biomas pyrolysis has shown that these kind of hybrid models are better at capturing experimental trends than purely data driven methods. Physically consistent predictions and lower data requirement for training have been realized from these models thanks to the incorporation of domain knowledge [7,21].

Physics-informed machine learning takes this a step further and, in addition to training with the constraint relationships that underpin first-principles models (mass and energy balance or reaction kinetics), one trains explicitly using physical laws. This way of thinking is becoming important in the context of combustion and thermochemical modeling, where the extrapolation from training domain has to be physically based for reliable predictions. The physics-informed models are robust in the face of limited data and also easier to interpret so that they can be used for real-time monitoring as well as control [18,40]. In biomolecular conversion research, machine learning approaches and thermochemical models can be integrated to create an interpretative model for understanding and optimizing the process [45]. The combination of machine learning and CFD in biomass gasification has also demonstrated enhanced prediction accuracy as well as computational efficiency in which re-action kinetics are embedded to the simulator [46]. Such a tight coupling between physics-based models and ML reduces the experimental need and increases generalization for complex, multiscale biofuel processes [47–48].

Most of the hybrid and physics-informed AI models in the literature are validated at lab, pilot plant, or simulation scale rather than against full industrial systems. Some papers show application to pilot plants, continuous reactors or engine test rigs that resemble industrial conditions and improve stability and generalization with respect to purely data-driven models. However, few large-scale industrial verification cases are available until now since many implementation aspects such as data availability and integration complexity or computational capacity constrain the validation of balancing mechanisms: most applications

are either pre-industrial or semi-industrial at present. Table 2 explains the Overview and Comparative Strengths of Machine Learning Approaches.

Table 2. Overview and Comparative Strengths of Machine Learning Approaches

Approach Category	Representative Methods	Typical Data Scale	Core Strengths	Key Limitations	Common Application Domains	Remarks from Reported Studies
Classical Regression Models	Linear Regression, PLS	Small to medium, structured datasets	Simple structure; high interpretability; useful for sensitivity analysis	Poor representation of nonlinear effects; limited predictive power	Preliminary yield estimation; variable screening	Commonly adopted as baseline models
Support Vector Methods	SVR (linear, RBF, polynomial)	Medium-sized datasets	Strong generalization; effective for nonlinear relationships; robust to noise	Computational cost increases with dataset size; kernel selection required	Biodiesel yield prediction; emission estimation	Stable performance with limited data
Tree-Based Models	Decision Tree, Random Forest	Mixed numerical and categorical data	Captures nonlinear interactions; interpretable feature importance	Single trees prone to overfitting; ensembles increase computation	Feedstock classification; combustion and pyrolysis modeling	Random Forest improves stability and accuracy
Gradient Boosting Models	XGBoost, LightGBM, CatBoost	Medium to large datasets	High prediction accuracy; strong nonlinear handling	Requires careful tuning; risk of overfitting	Emission prediction; product distribution modeling	Often outperforms ANN for NOx prediction
Shallow Neural Networks	ANN, MLP	Medium to large datasets	Learns complex nonlinear mappings; flexible architecture	Sensitive to data size; tuning and overfitting issues	Ester conversion; gasification and pyrolysis processes	Effective for thermochemical systems
Deep Learning Models	CNN, LSTM, MIMO networks	Large datasets or augmented data	Temporal and spectral learning; multi-output capability	High data demand; limited interpretability	Time-series emission modeling; fuel property prediction	Not always superior to simpler models
Hybrid & Physics-Informed Models	Hybrid ML–mechanistic, PINNs	Moderate data with physical constraints	Physically consistent predictions; improved extrapolation	Complex formulation; requires domain expertise	Pyrolysis kinetics; combustion soft sensors; reactor modeling	Improved robustness through physical constraints

The ML model selection is process dependent and related to the available data. Pyrolysis has pronounced nonlinear effects from feedstock properties and operating conditions, whose yields and product distribution can be predicted by ensemble tree models or neural networks. Structured operating condition-based engine combustion and emission studies could be suitably handled by SVR/boosting with their consistent generalization capability, while LSTM is more suitable for time-dependent emissions estimation. Hybrid or physics-informed frameworks in the form of gasification and combustion are to be preferred, with consideration of mass and energy balance constraints for robust prediction under scarce data. In general, studies have shown that the models' performance is process dependent with some common model–process combinations found but not one model which performs well for all biofuel application.

4. Model evaluation and benchmarks

For clarity, validation strategies used in biofuel yield and emission modelling can be grouped into distinct categories based on their purpose and data structure. Cross-validation methods, including k-fold and nested cross-validation, are primarily used to assess model stability and reduce optimistic bias during model selection, especially for small laboratory datasets. External testing involves training models on one feedstock, reactor configuration, or operating condition and evaluating them on an independent dataset, providing a realistic measure of generalization to unseen systems. Rolling-window evaluation is applied to time-series data, such as transient engine or continuous reactor operation, where models are trained on historical data and tested on sequential future windows to assess temporal robustness. This structured presentation helps distinguish the role of each validation approach and clarifies their relevance to different biofuel modelling scenarios.

4.1. Common metrics

Model performance in biofuel production and emission prediction is commonly assessed using regression or classification metrics, depending on the nature of the output. In regression tasks, the coefficient of determination (R^2) is frequently reported to indicate the proportion of output variance explained by the model. Root mean squared error is widely used because it reflects the average prediction error while assigning greater weight to large deviations. Mean absolute error provides a direct measure of average error magnitude, while mean absolute percentage error is useful for expressing errors in relative terms, especially when variables span different value ranges. Alongside these metrics, residual analysis and uncertainty assessment are important for evaluating model reliability and identifying regions of poor performance, particularly in systems with variable biomass composition and complex thermochemical behavior [5].

For classification tasks, such as fuel quality grading or operating range identification, accuracy, precision, recall, and F1-score are typically employed. Accuracy gives an overall measure of correct predictions, while precision and recall describe class-specific behavior. The F1-score is especially relevant for imbalanced datasets or when both false alarms and missed detections are critical. In multi-output prediction problems involving parameters such as yield, viscosity, density, or emission factors, relative error measures like average percentage error and root relative squared error are preferred because outputs differ in scale [1]. Consistent selection and reporting of performance metrics, such as normalized R^2 and RMSE, are necessary to allow meaningful comparison across studies [5]. For multi-output models, aggregated loss functions are often adopted, where individual errors are combined using weighted schemes or task-specific scaling to ensure that outputs with different units and ranges contribute appropriately during model training.

4.2. Validation practices

Rigorous validation is essential for reliable prediction of biofuel production and emission models. Recent studies frequently apply nested or double-nested cross-validation to clearly separate model selection from performance assessment, which reduces optimistic bias. External hold-out validation is also common, where models trained on one feedstock, operating condition, or reactor configuration are tested on independent datasets. This approach provides a realistic estimate of performance when models are applied to new biomass types or plant designs. In cases of limited data availability, leave-one-out cross-validation is used, though it becomes computationally demanding for large datasets. For time-series data, rolling window validation is recommended to properly assess temporal stability and predictive consistency. Across all validation strategies, strict prevention of data leakage is necessary to ensure unbiased evaluation [5].

Uncertainty quantification is receiving increasing attention in this field. Techniques such as prediction intervals, bootstrap resampling, and ensemble variance are employed to express confidence in model outputs and to identify unreliable predictions. These methods are particularly useful when training data are scarce or when operating conditions vary widely. Standard k-fold cross-validation remains a widely adopted approach

for robust performance estimation by repeatedly training and testing models across different data partitions [49–52]. Literature also reports that some models achieve high R^2 values on single-site datasets but show poor generalization when applied to different feedstocks or reactor systems. This observation highlights the need for large, multi-site datasets and blind validation strategies. Models trained on heterogeneous data are more likely to achieve strong generalization, which is essential for industrial deployment [1].

5. Applications and case studies (Summarized Examples)

5.1. Biodiesel transesterification

Several works show the ability of AI methods to reckon the methyl ester formation in a transesterification process. Artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS) and response surface methodology (RSM) coupled with machine learning have been extensively implemented. These models consider methanol-to-oil molar ratio, catalyst concentration, mixing rate and retention time as the major reaction parameters. They perform well in laboratory data sets for which they predict with high accuracy capturing nonlinear combinations of process variables poorly described by classical kinetic models. In some case studies, the models based on ANN and ANFIS were able to predict ester conversion of transesterification reaction with small errors and researchers could find near optimal operating point without performing many experimental trials.

5.2. Pyrolysis product distribution

Machine learning techniques have been used to estimate the product distribution of bio-oil, char and syngas in biomass pyrolysis. Typical approaches are random forests and deep neural networks that are capable of capturing complex interactions between biomass composition, reactor temperature, heating rate, particle size and vapor residence time. The results of feature importance analyses frequently identify temperature, heating rate and feedstock composition as the most important variables. This trend extends across multiple datasets, demonstrating the utility of ML for divining the nuanced process–property relationships which drive tune-ability in operation and feedstock selection. Deep architectures have also demonstrated potential for processing high-dimensional inputs signals such as spectroscopic signatures of biomass.

5.3. Engine combustion and emissions

Multilayer perceptron (MLP) and ensemble are employed to predict combustion outputs and emission contents including NOx, CO, HC, smoke opacity as well as particulate matter for studies on biodiesel blends and engine performance [26,27]. The models take many inputs, such as engine load and speed, fuel injection timing, exhaust gas temperature and the properties of the fuel. They are predictive over steady state and transient operating conditions. Some literature includes response surface methodology (DoE) to minimize the number of engine test while having its operability region widely covered. This DoE, ML hybrid approach is beneficial to find blend ratios that achieve a balance between the performance and emissions of the engine using biodiesel blends and presents a practical method for the analysis of utilisation of biodiesel in CI engines.

6. Challenges and limitations

6.1. Data scarcity and heterogeneity

The biofuels studies greatly depend on experimental data, which are generally small as well as collected over limited ranges of operation. There are still limited public benchmark datasets, that have precluded systematic comparison of modelling approaches. Several of the reported models are calibrated for laboratory-scale experiments with particular slurry or catalyst types or engine conditions in mind. In practice, closely associated biomass species, types of transesterification chemistry and reactor design/fuel modification approaches or engine designs may be missing in the data sets. Consequently, the models trained on one

feedstock or equipment package have low transferability, leading to a chasm between laboratory success and industrial performance. When data are sparse and diverse, the need for appropriate imputation or variable selection increases. Literature shows that imprecise treatment of missing values and outliers fosters overfitting and degradation in generalization across feedstocks/operating conditions.

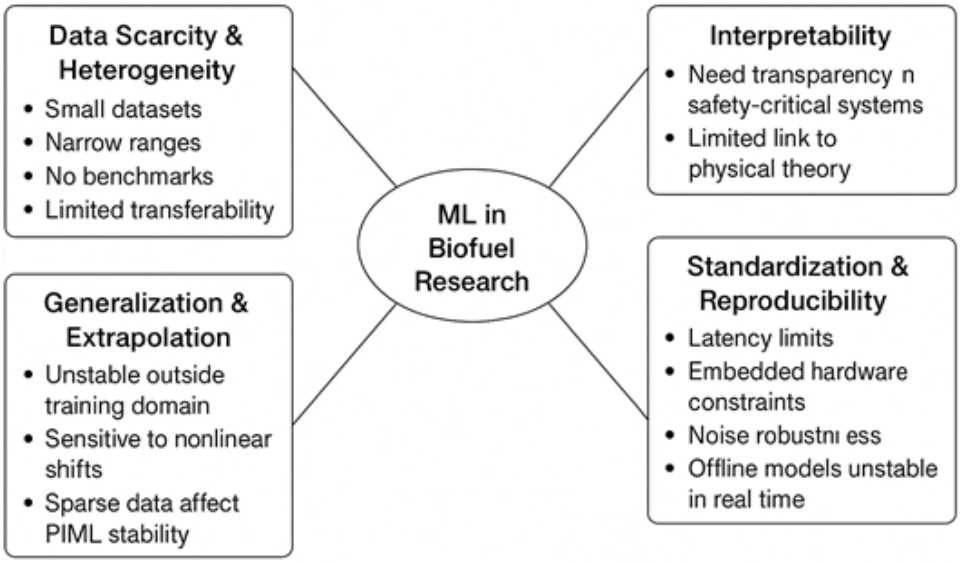

Figure 3. Key Challenges and Limitations in Applying Machine Learning to Biofuel Research

The elements in Figure 3 are included to visually consolidate the key limitations identified in current machine learning applications for biofuel research. Data scarcity highlights the dependence on small, single-site experimental datasets, which restricts model robustness. Limited generalization reflects the difficulty of applying trained models to new feedstocks, reactors, or operating ranges. The interpretability component emphasizes concerns related to black-box models, which limit physical insight and industrial trust. Issues related to standardization and real-time deployment point to practical barriers in transferring laboratory-scale models to operational systems. Together, these elements explain why gaps persist between reported model accuracy and real-world applicability.

6.2. Generalization and extrapolation

Data Analysing the experiments A common finding in all studies is that purely data driven models often behave erratically when they are presented with conditions outside the domain on which they have been trained. This is especially evident in pyrolysis and engine emission prediction, where small variations in temperature, or fuel composition can result in a nonlinear shift of an output. Hybrid methods and physics-informed ML (PIML) also serve to include constraints due to reaction kinetics, mass balances or combustion theory, thus enhancing stability. Nevertheless, good extrapolation is difficult to achieve in practice, particularly with relatively low numbers of samples or when the underlying physical processes are highly nonlinear and characterized by multi-step reactions. Besides, the chemical makeup of different biodiesels and the diverse operation conditions of engines greatly enlarges the difficulty in precisely predicting and generalizing [53-62]. Nevertheless, particular issues of anomalies to the accuracy prediction and potential overfitting in models such as XGBoost remain outstanding and require better optimization using methods such as feature selection, model tunning or coupling with DEEP learning in order to get significantly improved performance for practical scenarios [63-67].

6.3. Interpretability

Transparency models are often sought by industrial users, regulators, and process engineers who want to understand the reasons for a prediction. This is particularly evident in safety-critical domains such as combustion or reactor control. To interpret, you use tools such as ranking of feature-importance, partial dependence plots, values and surrogate models. Although these approaches are valuable, they often do not provide mechanistic descriptions that can relate back to the underlying chemistry, thermodynamics or flow physics. Closing this gap is an important open problem. For example, the black-box nature of many sophisticated machine learning models (such as deep neural networks), often restrict their use directly in industrial settings where understanding causality is a must for optimizing and troubleshooting processes [68-75]. This gap in interpretability compels one to trade off between prediction accuracy and model transparency, that is already an issue especially in domains that depend on validity and causal information [76-82].

6.4. Standardization and reproducibility

Discrepant preprocessing steps, model architectures, hyperparameter choices and validation procedures hinder the comparison with other studies. Some publications present only R^2 for one train–test split, whereas others employ cross-validation or external validation. Reproducibility is also hampered by differences in the recording of metrics, and patchy documentation of experimental conditions. This lack of standardization is a barrier to the creation of benchmarks that can be used with confidence as references to compare machine learning methods in biofuels research [83-95]. Developing common evaluation metrics and benchmark databases could also help to mitigate these problems, thus enabling engineers' research in fair and comparable grounds. Further, since the laboratory-scale data-trained ML models are heavily used at large scales with limited dataset or little margin to train these models, effective extrapolation techniques and solid uncertainty quantification is needed for stability and accuracy [96-102].

6.5. Real-time deployment constraints

For applications like engine control, process monitoring or soft sensing constraint-intensive even low-latency and -computing predictive models are essential. Models have to react fast, be robust against sensor noise and capable of real-time conditioning. A large number of deep learning models trained in simulation fail to retain robustness once they are transferred to embedded systems or control hardware. It remains a practical limitation to obtain high prediction accuracy while maintaining the real-time requirement, particularly for complex models that utilize extensive sets of features or time-series inputs. Such obstacles are frequently tackled by model quantization, pruning, and distillation approaches that reduce the complexity of DNN without a substantial degradation in linear classification (this allows their use in real time industrial systems [103-108]). Moreover, the elusive nature of complexity in the decision-making process of complex AI models (as those employed for prediction of biofuel yield and combustion emissions) may hamper their utility to support new scientific discoveries and discover new insights [109-116].

Limitations related to data scarcity and model generalization have been reported in several published studies and have direct implications for real-world application. For example, many pyrolysis yield models trained on single-feedstock laboratory datasets show high accuracy under controlled conditions but perform poorly when applied to different biomass types or reactor configurations, leading to large prediction errors during scale-up. Similar issues are observed in engine emission prediction, where models calibrated using steady-state test data fail to capture transient driving conditions, resulting in underestimation of NOx and particulate emissions in practical engine operation. Studies on biodiesel transesterification also report that models developed with narrow ranges of alcohol-to-oil ratios or catalyst concentrations exhibit limited transferability to industrial-scale reactors. These examples demonstrate that insufficient data diversity and weak external validation can restrict the reliability of machine learning models beyond the experimental domain.

7. Research gaps and recommended directions

7.1. Public, multi-site benchmark datasets

The development of common well documented data sets across variety feedstocks, equipment reactors types and engines platforms is by itself an important seed. Each data set should be accompanied with standardized metadata: sampling protocols, sensor calibration, units and experimental uncertainties) and versioning. Such resources would facilitate the making of reproducible comparisons, transfer learning, and prevent an overfitting to single-site specificities. Offering: Raw time series, derived physicochemical features, categorical labels for process configuration and clear provenance information.

7.2. Rigorous external validation protocols

Normalise external validation, rather than as an anomaly. Standard schemes should involve leavefeedstock- out and leave-site-out splits, temporal holdouts for sequential processes, as well as blind challenges where participants predict heldout datasets. Nesting cross-validation as well as hyper parameter tuning help diminishing optimistic bias. Reported results must provide calibration diagnostics and ranges of uncertainty, not just point estimates.

7.3. Uncertainty quantification and risk-aware reporting

Promote the standard inclusion of uncertainty estimates in model outputs. Quantify epistemic and aleatoric uncertainty with prediction intervals, Bayesian approximations, ensemble spread or conformal prediction. Report your current amount of uncertainty alongside the point predictions in papers and in any interface you deploy. This enables risk-conscious decision-making in the process control and regulatory environment.

Figure 4. Research Gaps and Future Directions in Machine Learning for Biofuel Modelling

Figure 4 is designed to map the identified challenges to concrete research directions. The focus on benchmark and multi-site datasets addresses data heterogeneity and reproducibility issues. Physics-aware and hybrid modelling approaches respond to generalization and extrapolation limitations. Uncertainty quantification and explainable models are included to support reliable decision-making and regulatory acceptance. Directions related to transfer learning, lifecycle integration, and real-time deployment reflect the need to move from isolated predictions toward scalable, system-level solutions. Each visual element thus represents a targeted response to limitations discussed earlier, guiding future research priorities.

7.4. Physics-aware hybrid modelling

Novel hybrid techniques that closely integrate low-dimensional mechanistic models with data-driven methods. Exploit ML to train residuals, estimate poorly known kinetic parameters or replace with data cheaper sub-models. Develop loss functions which punish deviations from conservative laws and/or impose the physics. ASTs should be evaluated for extrapolation capability against interpolation accuracy.

7.5. Domain-specific explainability

Create explain ability techniques that can map model behaviour to chemically/physically interpretable quantities. This encompasses rule extraction techniques that generate surrogate kinetic-like relationships, goal-oriented SHAP analyses focussed on thermodynamically feasible directions and visual analytics components that connect feature importance with reaction pathways or transport processes. These are the tools that will help to achieve regulator approval and engineer confidence.

7.6. Transfer learning and few-shot adaptation

Create workflows for adapting pertained models to new feedstocks or equipment using minimal new data. Explore meta-learning, domain-adaptation, and physics-guided fine-tuning to reduce experimental burden. Publicly release pertained model weights and documented fine-tuning recipes to accelerate adoption.

7.7. Lifecycle and multi-objective integration

Integrate ML predictions into broader sustainability assessments. Couple yield and emissions models with lifecycle assessment (LCA) and techno-economic analysis (TEA) to enable multi-objective optimization that balances environmental, economic, and operational goals. Provide standardized APIs and data formats so ML models can plug into LCA/TEA tools.

7.8. Standards for pre-processing, reporting, and reproducibility

Define community guidelines for pre-processing steps (missing-data handling, scaling, encoding), model reporting (hyper parameters, code, random seeds), and metric selection. Encourage publication of code, trained models, and minimal runnable examples. Standardized reporting will facilitate meta-analyses and accelerate progress.

7.9. Real-time and embedded deployment studies

Research practical constraints for deploying models in control loops: latency, robustness to sensor drift, explainable fail-safe modes, and model update strategies. Publish case studies that document hardware choices, quantifiable latency budgets, and resilience tests under noisy sensor inputs.

7.10. Cross-disciplinary testbeds and benchmarks

Establish collaborative testbeds that unite chemists, process engineers, machine learners, and regulators. Run community benchmark tasks with defined scoring rules (accuracy, calibration, computational cost, interpretability) to surface practical trade-offs and accelerate translation.

8. Conclusion

This review considers the state-of-the-art in applications of AI & ML to predict biofuel yield and combustion emissions throughout processes from biomass conversion to use in engines. The reviewed works show that the data-driven models, which include ensemble learning techniques, artificial neural networks, and physics-informed hybrid models are able to capture nonlinear interactions between feedstock properties, operating conditions and performance indicators. Under laboratory or pilot-scale conditions, these strategies generally demonstrate high predictive accuracy and less reliance on time-consuming experimental trials for accelerating process optimization and parameter screening However, some structural constraints are still apparent. The majority of the models cited here are trained on small datasets from a single reactor location and have limited operational ranges, making transferability across feedstocks, reactors, and engine platforms difficult. In addition, such reproducibility is impeded by differences in data preprocessing, featured engineering and validation setups that prevent fair comparison across studies. High reported R^2 values frequently indicate overfitting of the model to internal validation as opposed to actual generalization and models tend to perform poorly in applications outside their training scope or under dynamic industrial conditions. The review suggests that the next advances will come from improved modelling practice, rather than more sophisticated algorithms alone. The creation of multi-site benchmark datasets, utilization of stringent external validation, and the incorporation of uncertainty quantification will be necessary. Hybrid and physics-informed models have achieved notable results by incorporating mass, energy or kinetic constraints that can lead to better extrapolation and more interpretable outputs given limited data. Real-time deployment issues and standardized reporting guidelines will also aid practical application. With the integrated approach, AI prediction tools prove to be a good supplement for traditional models in the field of biofuels. With improved data sets and validation, such approaches can underpin dependable decisions for sustainable biofuel production and control of emissions.

Author Contributions

Pallavi Vishnu Kharat contributed to conceptualization, literature survey, data curation, and preparation of the original draft. **Beena Nawghare** provided support for conceptual framing, domain-specific inputs, and critical review of the manuscript. **N. Alangudi Balaji** contributed to methodology formulation, analysis of machine learning techniques, and technical validation. **Vishvas V. Kalunge** assisted with data interpretation, comparative analysis, and manuscript editing. **Charu P. Kumbhare** contributed to process-level insights, application-oriented discussion, and validation of technical content. **Tejasvini Rahul Katkar** supported data organization, figure preparation, and formatting. **Sagar Arjun Dalvi and Shital Yashwant Waware** contributed to literature synthesis, reference management, and consistency checks across sections. **Anant Sidhappa Kurhade** provided overall supervision, research guidance, critical revision, and final approval of the manuscript. All authors reviewed the manuscript and approved the final version for submission.

Acknowledgments

The authors would like to express their sincere gratitude to Dr. D. Y. Patil Institute of Technology and DPGU, School of Technology and Research - Dr. D. Y. Patil Unitech Society, Pimpri, Pune, for providing the necessary support and research infrastructure.

Conflict of interest

The authors declare no conflict of interest

References

1. Jeon PR, Moon JH, Ogunsola NO, Lee SH, Ling JLJ, You S, Park YK. Recent advances and future prospects of thermochemical biofuel conversion processes with machine learning. Chemical Engineering Journal. 2023;471:144503. doi:10.1016/j.cej.2023.144503
2. Nguyen TH, Paramasivam P, Dong VH, Le HC, Nguyen DC. Harnessing a better future: exploring AI and ML applications in renewable energy. JOIV International Journal on Informatics Visualization. 2024;8(1):55. doi:10.62527/joiv.8.1.2637
3. Wang Z, Peng X, Xia A, Shah AA, Huang Y, Zhu X, Zhu X, Liao Q. The role of machine learning to boost the bioenergy and biofuels conversion. Bioresource Technology. 2021;343:126099. doi:10.1016/j.biortech.2021.126099
4. Liao M, Yao Y. Applications of artificial intelligence-based modeling for bioenergy systems: a review. GCB Bioenergy. 2021;13(5):774. doi:10.1111/gcbb.12816
5. Ascher S, Watson I, You S. Machine learning methods for modelling the gasification and pyrolysis of biomass and waste. Renewable and Sustainable Energy Reviews. 2021;155:111902. doi:10.1016/j.rser.2021.111902
6. Le TT, Paramasivam P, Adril E, Quý NV, Le MX, Duong MT, Le HC, Nguyen AQ. Unlocking renewable energy potential: harnessing machine learning and intelligent algorithms. International Journal of Renewable Energy Development. 2024;13(4):783. doi:10.61435/ijred.2024.60387
7. Lee H, Choi IH, Hwang KR. A comprehensive review of machine learning prediction in the production of bio-oil from lignocellulose via pyrolysis. Research Square. 2024. doi:10.21203/rs.3.rs-3830648/v1
8. Li F, Li Y, Novoselov KS, Liang F, Meng J, Ho S, et al. Bioresource upgrade for sustainable energy, environment, and biomedicine. Nano-Micro Letters. 2023;15(1). doi:10.1007/s40820-022-00993-4
9. Li H, Chen J, Zhang W, Zhan H, He C, Yang Z, Peng H, Leng L. Machine-learning-aided thermochemical treatment of biomass: a review. Biofuel Research Journal. 2023;10(1):1786. doi:10.18331/brj2023.10.1.4
10. Yelgel ÖC, Yelgel C. The role of machine learning methods for renewable energy forecasting. IntechOpen eBooks. 2024. doi:10.5772/intechopen.1007556
11. Clauser NM, Felissia FE, Área MC, Vallejos ME. Integrating the new age of bioeconomy and Industry 4.0 into biorefinery process design. BioResources. 2022;17(3):5510. doi:10.15376/biores.17.3.clauser
12. Nair LG, Verma P. Harnessing carbon potential of lignocellulosic biomass: advances in pretreatments, applications, and the transformative role of machine learning in biorefineries. Bioresources and Bioprocessing. 2025;12(1). doi:10.1186/s40643-025-00935-z
13. Gil MV, Jablonka KM, García S, Pevida C, Smit B. Biomass to energy: a machine learning model for optimum gasification pathways. Digital Discovery. 2023;2(4):929. doi:10.1039/d3dd00079f
14. Qing-sheng X, Du L, Deng R. Using machine learning to predict biochar yield and carbon content. BioResources. 2024;19(3):6545. doi:10.15376/biores.19.3.6545-6558
15. Li J, Chen Y, Wang C, Chen H, Gao Y, Meng J, et al. Optimizing biochar for carbon sequestration using machine learning and NLP. Biochar. 2025;7(1). doi:10.1007/s42773-024-00424-0
16. Ali MI, Shabbir K, Ali S, Mohsin M, Kumar A, Aziz M, et al. How artificial intelligence has revolutionized biotechnology. Nepal Journal of Biotechnology. 2024;12(1). doi:10.54796/njb.v12i1.312
17. Kant G, Hasan A, Mani P, Pandey A, Srivastava S. The generational shift in biofuels. Biomass and Bioenergy. 2025;196:107757. doi:10.1016/j.biombioe.2025.107757
18. Osman AI, Nasr M, Farghali M, Rashwan AK, Abdelkader A, Al-Muhtaseb AH, et al. Optimizing biodiesel production from waste with ML. Environmental Chemistry Letters. 2024;22(3):1005. doi:10.1007/s10311-024-01700-y
19. Meena MK, Shubham S, Paritosh K, Pareek N, Vivekanand V. Predicting energy from biomass using AI. Bioresource Technology. 2021;340:125642. doi:10.1016/j.biortech.2021.125642
20. Rahimi M, Mashhadimoslem H, Thanh HV, Ranjbar B, Khosrowshahi MS, Rohani A, Elkamel A. Yield prediction of biomass-based products by ML schemes. Energy. 2023;283:128546. doi:10.1016/j.energy.2023.128546
21. Ramalingam K, Abdullah MZ, Elumalai PV, Xu Y, Prasad KSN, Chan CK, et al. Maximizing biofuel production using ML. Scientific Reports. 2025;15(1). doi:10.1038/s41598-025-18757-6
22. Anbarasu K, Sundaram T, Sathishkumar K, Alam MM, Al-Sehemi AG, Devarajan Y. Harnessing artificial intelligence for sustainable bioenergy. Bioresource Technology. 2024;131893. doi:10.1016/j.biortech.2024.131893
23. Arif M, Alalawy AI, Zheng Y, Koutb M, Kareri T, Salama E, Li X. AI and ML models for biodiesel optimization. Sustainable Energy Technologies and Assessments. 2024;73:104097. doi:10.1016/j.seta.2024.104097
24. Nath S. Biotechnology and biofuels. Discover Energy. 2024;4(1). doi:10.1007/s43937-024-00032-w
25. Aghbashlo M, Peng W, Tabatabaei M, Kalogirou SA, Soltanian S, Hosseinzadeh-Bandbafha H, et al. Machine learning in biodiesel research. Progress in Energy and Combustion Science. 2021;85:100904. doi:10.1016/j.pecs.2021.100904
26. Awogbemi O, Kallon DVV. ML technologies in biodiesel production. Frontiers in Energy Research. 2023;11. doi:10.3389/fenrg.2023.1122638

27. Nneoma UC, Chukwudi OF, Nnenna UJ, Ugwu OPC. Hybrid biofactories for enhanced biofuel production. Frontiers in Energy Research. 2025;13. doi:10.3389/fenrg.2025.1654079
28. Abdullah M, Malik HA, Ali AM, Boopathy R, Vo HNP, Danaee S, et al. AI-driven algae biorefineries. Current Pollution Reports. 2025;11(1). doi:10.1007/s40726-025-00352-y
29. Geng Y, Shaukat A, Azhar W, Raza QUA, Tahir A, Abideen MZU, et al. Microalgal biorefineries: systematic review. Biotechnology for Biofuels and Bioproducts. 2025;18(1). doi:10.1186/s13068-025-02694-7
30. Çokkızgın H, Çokkızgın A, Girgel Ü. Reducing carbon footprint of field crops. Applied Ecology and Environmental Research. 2025;23(4):7641. doi:10.15666/aeer/2304_76417661
31. Ahmad I, Sana A, Kano M, Cheema II, Menezes BC, Shahzad J, et al. Machine learning applications in biofuels' life cycle. Energies. 2021;14(16):5072. doi:10.3390/en14165072
32. Wang C. Biomass fluidized bed gasification model with ML and CFD. arXiv. 2025. doi:10.48550/arXiv.2509.06056
33. Costa M, Piazzullo D. Biofuel powering of ICE engines: production and CFD modeling. Frontiers in Mechanical Engineering. 2018;4. doi:10.3389/fmech.2018.00009
34. Arumugam S, Muthaiyan R, Dhairiyasamy R, Rajendran S. Enhancing biodiesel stability and performance. Discover Chemical Engineering. 2024;4(1). doi:10.1007/s43938-024-00041-0
35. Hong B, Burghout W. Adaptive estimation in NOx emission modelling. Emission Control Science and Technology. 2024;10(2):93. doi:10.1007/s40825-024-00241-8
36. Jiang P, Wang C, She WJ, Ye W, Li Y chen, Ji T, et al. ML-accelerated process modelling for biomass conversion. Renewable and Sustainable Energy Reviews. 2025;226:116439. doi:10.1016/j.rser.2025.116439
37. Coniglio L, Bennadji H, Glaude P, Herbinet O, Billaud F. Combustion chemical kinetics of biodiesel. Progress in Energy and Combustion Science. 2013;39(4):340. doi:10.1016/j.pecs.2013.03.002
38. Jana S, Chatterjee D, Pal N, Pal K, Roy K, Bashak S. AI in soil health monitoring. IJRASET. 2024;12(10):1327. doi:10.22214/ijraset.2024.64871
39. Mahanty B, Gharami M, Haldar D. ML modelling for ionic liquid-aided biomass pretreatment. BioEnergy Research. 2024;17(3):1569. doi:10.1007/s12155-024-10747-2
40. Zhu L, Chen X, Ouyang B, Yan W, Lei H, Chen Z, Luo Z. ML for hydrodynamics and reactions in multiphase flows. Industrial & Engineering Chemistry Research. 2022;61(28):9901. doi:10.1021/acs.iecr.2c01036
41. Ge X, Zhang T, Mukherjee S, Chen Y, Wang X, Chen N, et al. Optimizing hydrothermal treatment with ML. Biochar. 2025;7(1). doi:10.1007/s42773-025-00485-9
42. Mahanty B, Gharami M, Haldar D. ML models for ionic liquid biomass pretreatment. Research Square. 2024. doi:10.21203/rs.3.rs-3916542/v1
43. Nachippan M, Pathmanabhan P, Nagappan B, Upadhye VJ, Kaliappan N, Balaji V, Priya K. ML predictions for engine performance with nano-additives. Scientific Reports. 2025;15(1). doi:10.1038/s41598-025-02388-y
44. Brusa A, Giovannardi E, Barichello M, Cavina N. Data-driven surrogate models for NOx emissions. Energies. 2022;15(21):8088. doi:10.3390/en15218088
45. Khandelwal K, Dalai AK. Predicting gas yields of SCWG using ML. Molecules. 2024;29(10):2337. doi:10.3390/molecules29102337
46. Wang C. Biomass gasification ML–CFD model. arXiv. 2025. doi:10.48550/arXiv.2509.06056
47. Caprio UD, Leblebici ME. SINDybrid: automatic generation of hybrid models. arXiv. 2025. doi:10.48550/arXiv.2506.12498
48. Sharma N, Liu YA. Hybrid science-guided ML for chemical processes. arXiv. 2021. doi:10.48550/arXiv.2112.01475
49. Driessen JLSP. Integration of biomass conversion, cell factory optimization and bioprocess design. Research Portal Denmark. 2022.
50. Spooner MP. Statistical data analysis tools for bio-based manufacturing. Research Portal Denmark. 2018.
51. Roy S, Khekare G, Chhajed S, Victor A. ML for biochar safety and pollutant removal. Frontiers in Soil Science. 2025;5. doi:10.3389/fsoil.2025.1661097
52. Shahbeik H, Shafizadeh A, Nadian MH, Jeddi D, Mirjalili S, Yang Y, et al. Evolutionary ML for co-pyrolysis optimisation. Journal of Cleaner Production. 2023;387:135881. doi:10.1016/j.jclepro.2023.135881
53. Öner İV. ML algorithms applied to internal combustion engine studies. DergiPark. 2025.
54. Liu D, Sun N. Prospects of AI in sustainable separation processes. Frontiers in Sustainability. 2023;4. doi:10.3389/frsus.2023.1210209
55. Jacobsen C. Enhancing physical modelling with physics-aware ML. University of Michigan Deep Blue. 2024.
56. Gupta R, Ouderji ZH, Uzma U, Yu Z, Sloan WT, You S. Machine learning for sustainable organic waste treatment. npj Materials Sustainability. 2024;2(1). doi:10.1038/s44296-024-00009-9
57. Ojadi JO, Owulade OA, Odionu CS, Onukwulu EC. Deep learning models for predicting industrial environmental impact. IJSRSET. 2025;12(2):119. doi:10.32628/ijsrset25122109
58. Valizadeh A, Amirhosseini MH. Machine learning in lithium-ion batteries. SN Computer Science. 2024;5(6). doi:10.1007/s42979-024-03046-2

59. Kurhade AS, Gadekar T, Siraskar GD, Jawalkar SS, Biradar R, Kadam AA, Yadav RS, Dalvi SA, Waware SY, Mali CN. Thermal performance analysis of electronic components on different substrate materials. J Mines Met Fuels. 2024 Oct 1;72(10). https://doi.org/10.18311/jmmf/2024/45569
60. Kurhade AS, Siraskar GD, Jawalkar SS, Gadekar T, Bhambare PS, Biradar R, Yadav RS, Waware SY, Mali CN. The impact of circular holes in twisted tape inserts on forced convection heat transfer. J Mines Met Fuels. 2024 Oct 16;72(9):1005-12. https://doi.org/10.18311/jmmf/2024/45505
61. Kurhade AS, Bhambare PS, Desai VP, Murali G, Yadav RS, Patil P, Gadekar T, Biradar R, Kirpekar S, Charwad GA, Waware SY. Investigating the effect of heat transfer influenced by wavy corrugated twisted tape inserts in double pipe heat exchangers. J Adv Res Fluid Mech Therm Sci. 2024;122:146-55. https://doi.org/10.37934/arfmts.122.2.146155
62. Kurhade AS, Murali G, Jadhav PA, Bhambare PS, Waware SY, Gadekar T, Yadav RS, Biradar R, Patil P. Performance analysis of corrugated twisted tape inserts for heat transfer augmentation. J Adv Res Fluid Mech Therm Sci. 2024;121(2):192-200. https://doi.org/10.37934/arfmts.121.2.192200
63. Yadav RS, Nimbalkar A, Gadekar T, Patil P, Patil VN, Gholap AB, Kurhade AS, Dhumal JR, Waware SY. Comparison of experimental and numerical investigation of mono-composite and metal leaf spring. J Mines Met Fuels. 2024 Aug 1;72(8). https://doi.org/10.18311/jmmf/2024/45325
64. Kurhade AS, Warke P, Maniyar K, Bhambare PS, Waware SY, Deshpande S, Harsur S, Ingle M, Kolhe P, Patil PA, Jadhav P. Wind rose analysis of temperature variation with sensor implantation technique for wind turbine. J Adv Res Fluid Mech Therm Sci. 2024;122(1):1-8. https://doi.org/10.37934/arfmts.122.1.118
65. Kurhade AS, Siraskar GD, Bhambare PS, Kaithari DK, Dixit SM, Waware SY. Enhancing smartphone circuit cooling: a computational study of PCM integration. J Adv Res Numer Heat Trans. 2024 Nov 30;27(1):132-45. https://doi.org/10.37934/arnht.27.1.132145
66. Kurhade AS, Darade MM, Siraskar GD, Biradar R, Mahajan RG, Kardile CS, Waware SY, Yadav RS. State-of-the-art cooling solutions for electronic devices operating in harsh conditions. J Mines Met Fuels. 2024 Aug 1;72(8). https://doi.org/10.18311/jmmf/2024/45374
67. Yadav RS, Gadekar T, Gundage V, Patil P, Patil A, Patil P, Patil A, Sutar R, Kurhade AS. Numerical and experimental investigation of the effect of overlapping angle on strength and deformation of curved plate joined using arc welding process. J Mines Met Fuels. 2024 Oct 1;72(10). https://doi.org/10.18311/jmmf/2024/45697
68. Kurhade AS, Bhambare PS, Siraskar GD, Dixit SM, Purandare PS, Waware SY. Computational study on thermal management of IC chips with phase change materials. J Adv Res Numer Heat Trans. 2024;26(1):34-43. https://doi.org/10.37934/arnht.26.1.3443
69. Yadav RS, Gandhi P, Veeranjaneyulu K, Gaji R, Kirpekar S, Pawar D, Khairnar YS, Patil S, Kurhade AS, Patil SP. Influence of plate thickness on the mechanical behaviour of mild steel curved plates: an experimental study. J Mines Met Fuels. 2024 Dec 1;72(12). https://doi.org/10.18311/jmmf/2024/46253
70. Chippalkatti S, Chekuri RB, Ohol SS, Shinde NM, Barmavatu P, Shelkande VD, Murali G, Kurhade AS. Enhancing heat transfer in micro-channel heat sinks through geometrical optimization. J Mines Met Fuels. 2025 Mar 1;73(3). https://doi.org/10.18311/jmmf/2025/47773
71. Raut PN, Dolas AS, Chougule SM, Darade MM, Murali G, Waware SY, Kurhade AS. Green adsorbents for heavy metal removal: a study on zinc ion uptake by Tinospora cordifolia biocarbon. J Mines Met Fuels. 2025 Jan 1;73(1). https://doi.org/10.18311/jmmf/2025/47121
72. Kurhade AS, Siraskar GD, Bhambare PS, Murali G, Deshpande SV, Warke PS, Waware SY. Simulation and analysis of heat transfer in counter-flow helical double-pipe heat exchangers using CFD. Int J Mod Phys C. 2025 Mar 15. https://doi.org/10.1142/S0129183125500433
73. Patil Y, Tatiya M, Dharmadhikari DD, Shahakar M, Patil SK, Mahajan RG, Kurhade AS. The role of AI in reducing environmental impact in the mining sector. J Mines Met Fuels. 2025 May 1;73(5).
74. Waware SY, Ahire PP, Napate K, Biradar R, Patil SP, Kore SS, Kurhade AS. Advancements in heat transfer enhancement using perforated twisted tapes: a comprehensive review. J Mines Met Fuels. 2025 May 1;73(5). https://doi.org/10.18311/jmmf/2025/48438
75. Chougule SM, Murali G, Kurhade AS. Finite element analysis and design optimization of a paddle mixer shaft. J Mines Met Fuels. 2025 May 1;73(5). https://doi.org/10.18311/jmmf/2025/48664
76. Chougule SM, Murali G, Kurhade AS. Failure investigation of the driving shaft in an industrial paddle mixer. J Mines Met Fuels. 2025 May 1;73(5). https://doi.org/10.18311/jmmf/2025/48627
77. Kurhade AS, Sugumaran S, Kolhalkar NR, Karad MM, Mahajan RG, Shinde NM, Dalvi SA, Waware SY. Thermal management of mobile devices via PCM. J Mines Met Fuels. 2025 May 1;73(5):1313-20. https://doi.org/10.18311/jmmf/2025/48437
78. Kurhade AS, Bhavani P, Patil SA, Kolhalkar NR, Chalapathi KS, Patil PA, Waware SY. Mitigating environmental impact: a study on the performance and emissions of a diesel engine fueled with biodiesel blend. J Mines Met Fuels. 2025 Apr 1;73(4):981-9. https://doi.org/10.18311/jmmf/2025/47669
79. Kurhade AS, Siraskar GD, Chekuri RB, Murali G, Pawar P, Patil AR, Waware SY, Yadav RS. Biodiesel blends: a sustainable solution for diesel engine performance improvement. J Mines Met Fuels. 2025 Mar 1;73(3). https://doi.org/10.18311/jmmf/2025/47628

80. Kurhade AS, Siraskar GD, Darade MM, Murali G, Katkar TR, Patil SP, Charwad GA, Waware SY, Yadav RS. Enhancement in heat transfer with nanofluids in double-pipe heat exchangers. J Mines Met Fuels. 2025 Jan 7;73(1):165-72. https://doi.org/10.18311/jmmf/2025/47225
81. Napte K, Kondhalkar GE, Patil SV, Kharat PV, Banarase SM, Kurhade AS, Waware SY. Recent advances in sustainable concrete and steel alternatives for marine infrastructure. Sustain Mar Struct. 2025 Jun 4:107-31. https://doi.org/10.36956/sms.v7i2.2072
82. Kurhade AS, Chougule SM, Kharat PV, Kondhalkar GE, Murali G, Raut PN, Charwad GA, Waware SY, Yadav RS. Integrated approach to enhance vehicle safety: a novel bumper design with energy-absorbing mechanisms. J Mines Met Fuels. 2025 Jan 1;73(1). https://doi.org/10.18311/jmmf/2025/47168
83. Yadav R, Nimbalkar A, Kirpekar S, Patil PJ, Dalvi SA, Jadhav PA, Kurhade AS, Wakchaure GN. Effect of transformed-induced plasticity steel plate thickness on ultimate tensile strength of butt welded joint using Nd:YAG laser. Int J Veh Struct Syst. 2024;16(6):857-62. https://doi.org/10.4273/ijvss.16.6.08
84. Deshpande SV, Pawar RS, Keche AJ, Kurhade A. Real-time surface finish measurement of stepped holding shaft by automatic system. J Adv Manuf Syst. 2025 Feb 25:1-26.
85. Ramani P, Reji V, Sathish Kumar V, Murali G, Kurhade AS. Deep learning-based detection and classification of moss and crack damage in rock structures for geo-mechanical preservation. J Mines Met Fuels. 2025 Mar 1;73(3). https://doi.org/10.18311/jmmf/2025/47760
86. Kurhade AS, Siraskar GD, Deshmukh MT, Patil PA, Chaudhari SS, Kadam AA, Dolas AS, Mahajan RG, Waware SY, Yadav RS. Impact of PCM on heat dissipation from IC chips. J Mines Met Fuels. 2025 Mar 1;73(3). https://doi.org/10.18311/jmmf/2025/47522
87. Kurhade AS, Kharat PV, Chougule SM, Darade MM, Karad MM, Murali G, Charwad GA, Waware SY, Yadav RS. Harnessing the power of plastic waste: a sustainable approach to fuel production. J Mines Met Fuels. 2025 Feb 1;73(2). https://doi.org/10.18311/jmmf/2025/47354
88. Sarode GC, Gholap P, Pathak KR, Vali PSNM, Saharkar U, Murali G, Kurhade AS. Edge AI and explainable models for real-time decision-making in ocean renewable energy systems. Sustain Mar Struct. 2025 Jun 24;7(3):17-42. https://doi.org/10.36956/sms.v7i3.2239
89. Chougule SM, Murali G, Kurhade AS. Dynamic simulation and performance evaluation of vibratory bowl feeders integrated with paddle shaft mechanisms. Adv Sci Technol Res J. 2025;19(7). https://doi.org/10.12913/22998624/203873
90. Chougule SM, Murali G, Kurhade AS. Design and analysis of industrial material handling systems using FEA and dynamic simulation techniques. J Sci Ind Res. 2025 Jun 18;84(6):645-53. https://doi.org/10.56042/jsir.v84i6.17512
91. Kurhade AS, Siraskar GD, Raut PN, Dolas AS, Murali G, Dalvi SA, Waware SY, Yadav RS. Investigating the impact of oxygenated additives on exhaust emissions from unleaded gasoline vehicles. J Mines Met Fuels. 2025 Feb 1;73(2). https://doi.org/10.18311/jmmf/2025/47410
92. Siraskar GD, Kurhade AS, Murali G, Prakash MA, Bharathiraja N, Dharmadhikari DD, Waware SY. Turbulence model comparison and optimal geometry identification in trapped vortex combustors: a RANS-based study. Int J Mod Phys C. 2025 Sep 24:2650020. https://doi.org/10.1142/S0129183126500208
93. Keloth Kaithari D, Kaulage A, Ayyappadas MT, Gholap P, Puri A, Bhandari MA, et al. A review of smart AI systems for real-time monitoring and optimization of ocean-based carbon capture, utilization, and storage networks. Appl Chem Eng. 2025 Sep 17;8(3):ACE-5747. https://doi.org/10.59429/ace.v8i3.5747
94. Dhamdhere P, Dixit SM, Tatiya M, Shinde BA, Deone J, Kaulage A, et al. AI-based monitoring and management in smart aquaculture for ocean fish farming systems. Appl Chem Eng. 2025 Sep 17;8(3):ACE-5746. https://doi.org/10.59429/ace.v8i3.5746
95. Keloth Kaithari D, Ayyappadas MT, Goel S, Shahin A, Patil SK, Chaudhari SS, et al. A review on GA-NN based control strategies for floating solar-ocean hybrid energy platforms. Appl Chem Eng. 2025 Sep 15;8(3):ACE-5745. https://doi.org/10.59429/ace.v8i3.5745
96. Bhambare PS, Kaulage A, Darade MM, Murali G, Dixit SM, Vali PSNM, et al. Artificial intelligence for sustainable environmental management in the mining sector: a review. Appl Chem Eng. 2025 Sep 18;8(3):ACE-5756. https://doi.org/10.59429/ace.v8i3.5756
97. Dharmadhikari DD, Ray A, Shinde BA, Raut SV, Taware RD, Desai S, et al. Machine learning applications in ore grade estimation and blending optimization for modern mining. Appl Chem Eng. 2025 Nov 6;8(4):ACE-5790. https://doi.org/10.59429/ace.v8i4.5790
98. Tatiya M, Darade MM, Shinde BA, Kumbhare MP, Taware RD, Chougule SM, et al. AI applications in tailings and waste management: improving safety, recycling, and water utilization. Appl Chem Eng. 2025 Nov 5;8(4):ACE-5789. https://doi.org/10.59429/ace.v8i4.5789
99. Upadhe SN, Mhamane SC, Kurhade AS, Bapat PV, Dhavale DB, Kore LJ. Water saving and hygienic faucet for public places in developing countries. In: Techno-Societal 2018: Proceedings of the 2nd International Conference on Advanced Technologies for Societal Applications. Vol 1. Cham: Springer; 2019. p. 617-24. https://doi.org/10.1007/978-3-030-16848-3_56

100. Kurhade AS, Siraskar GD, Darade MM, Dhumal JR, Kardile CS, Biradar R, Patil SP, Waware SY. Predicting heat transfer enhancement with twisted tape inserts using fuzzy logic techniques in heat exchangers. J Mines Met Fuels. 2024;72(7):743-50. https://doi.org/10.18311/jmmf/2024/45348
101. Kurhade AS, Siraskar GD, Kondhalkar GE, Darade MM, Yadav RS, Biradar R, Waware SY, Charwad GA. Optimizing aerofoil design: a comprehensive analysis of aerodynamic efficiency through CFD simulations and wind tunnel experiments. J Mines Met Fuels. 2024;72(7):713-24. https://doi.org/10.18311/jmmf/2024/45361
102. Kurhade AS, Kadam AA, Biradar R, Bhambare PS, Gadekar T, Patil P, Yadav RS, Waware SY. Experimental investigation of heat transfer from symmetric and asymmetric IC chips mounted on the SMPS board with and without PCM. J Adv Res Fluid Mech Therm Sci. 2024;121(1):137-47. https://doi.org/10.37934/arfmts.121.1.137147
103. Kurhade AS, Siraskar GD, Bhambare PS, Dixit SM, Waware SY. Numerical investigation on the influence of substrate board thermal conductivity on electronic component temperature regulation. J Adv Res Numer Heat Trans. 2024;23(1):28-37. https://doi.org/10.37934/arnht.23.1.2837
104. Kurhade AS, Waware SY, Munde KH, Biradar R, Yadav RS, Patil P, Patil VN, Dalvi SA. Performance of solar collector using recycled aluminum cans for drying. J Mines Met Fuels. 2024 May 1;72(5). https://doi.org/10.18311/jmmf/2024/44643
105. Kurhade AS, Kardekar NB, Bhambare PS, Waware SY, Yadav RS, Pawar P, Kirpekar S. A comprehensive review of electronic cooling technologies in harsh field environments: obstacles, progress, and prospects. J Mines Met Fuels. 2024;72(6):557-79. https://doi.org/10.18311/jmmf/2024/45212
106. Kurhade AS, Waware SY, Bhambare PS, Biradar R, Yadav RS, Patil VN. A comprehensive study on Calophyllum inophyllum biodiesel and dimethyl carbonate blends: performance optimization and emission control in diesel engines. J Mines Met Fuels. 2024;72(5):499-507. https://doi.org/10.18311/jmmf/2024/45188
107. Kurhade AS, Biradar R, Yadav RS, Patil P, Kardekar NB, Waware SY, Munde KH, Nimbalkar AG, Murali G. Predictive placement of IC chips using ANN-GA approach for efficient thermal cooling. J Adv Res Fluid Mech Therm Sci. 2024;118(2):137-47. https://doi.org/10.37934/arfmts.118.2.137147
108. Waware SY, Chougule SM, Yadav RS, Biradar R, Patil P, Munde KH, Kardekar NB, Nimbalkar AG, Kurhade AS, Murali G, Kore SS. A comprehensive evaluation of recent studies investigating nanofluids utilization in heat exchangers. J Adv Res Fluid Mech Therm Sci. 2024;119(2):160-72. https://doi.org/10.37934/arfmts.119.2.160172
109. Kurhade AS, Murali G, Rao TV. CFD approach for thermal management to enhance the reliability of IC chips. Int J Eng Trends Technol. 2022;71(3):65-72. https://doi.org/10.14445/22315381/IJETT-V71I3P208
110. Kurhade AS, Rao TV, Mathew VK, Patil NG. Effect of thermal conductivity of substrate board for temperature control of electronic components: a numerical study. Int J Mod Phys C. 2021 Oct 26;32(10):2150132. https://doi.org/10.1142/S0129183121501321
111. Waware SY, Kore SS, Kurhade AS, Patil SP. Innovative heat transfer enhancement in tubular heat exchanger: an experimental investigation with minijet impingement. J Adv Res Fluid Mech Therm Sci. 2024;116(2):51-8. https://doi.org/10.37934/arfmts.116.2.5158
112. Kurhade AS, Murali G. Thermal control of IC chips using phase change material: a CFD investigation. Int J Mod Phys C. 2022 Dec 28;33(12):2250159. https://doi.org/10.1142/S0129183122501595
113. Rami Reddy S, Murali G, Dhana Raju V. Assessment of diethyl ether as a fuel additive on diesel engine characteristics powered with waste mango seed biodiesel blend. Int J Ambient Energy. 2022 Dec 31;43(1):3365-76. https://doi.org/10.1080/01430750.2020.1824944
114. Emeema J, Murali G, Reddi BV, Mangesh VL. Investigations on paraffin wax/CQD composite phase change material: improved latent heat and thermal stability. J Energy Storage. 2024 Apr 30;85:111056. https://doi.org/10.1016/j.est.2024.111056
115. Rami Reddy S, Murali G, Dhana Raju V. Influence of decanol as fuel additive on characteristics of diesel engine powered with mango seed biodiesel blend. Int J Ambient Energy. 2022 Dec 31;43(1):2875-88. https://doi.org/10.1080/01430750.2020.1783356
116. Tamiloli N, Venkatesan J, Murali G, Kodali SP, Sampath Kumar T, Arunkumar MP. Optimization of end milling on Al-SiC-fly ash metal matrix composite using TOPSIS and fuzzy logic. SN Appl Sci. 2019;1(10):1204. https://doi.org/10.1007/s42452-019-1191-z