Fuel 374 (2024) 132396

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Full Length Article

Investigation of sawdust microwave-assisted pyrolysis by machine learning, Part I: Optimization insights by large language models

Bin Chen [a,e,1,*], Haoyu Wang [a,b,1], Xihe Qiu [b,1], Zilong Yin [a,b], Hangling Sun [c], Anji Li [d]

[a] University of Shanghai for Science and Technology, Shanghai, China
[b] School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China
[c] Hengtu Imalligent Technology (Shanghai) Co., Ltd., Shanghai, China
[d] Abbott Laboratories(Shanghai) Co., Ltd., Shanghai, China
[e] School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Keywords:
Bio-oil production
Catalytic pyrolysis
Sawdust
Large language model
Machine learning

ABSTRACT

Microwave-assisted catalytic pyrolysis shows promise for efficiently converting waste biomass, such as sawdust, into valuable bio-oil. However, current research on pyrolysis characteristics predominantly relies on conventional trial-and-error experimentation. This work pioneers the first use of large language models (LLMs) to gain insights into optimizing bio-oil yield by analyzing the effects of catalyst loading and pretreatment temperature on product distribution. Additionally, we encode the textual LLM outputs into distributed vectors via Word2Vec and concatenate them with artificial neural network (ANN) embeddings, capitalizing on the complementary strengths of data-driven and language models. Our results demonstrate superior accuracy and interpretability, providing better optimization insights on heating dynamics and energy efficiency while avoiding extensive experimental costs. This research establishes the prospects of LLMs as supplementary tools for thermochemical conversion studies, potentially reducing resource-intensive lab trials through reliable data-driven predictions.

1. Introduction

With the growing demand for renewable fuels [1], there has been increasing interest in converting biomass waste into value-added chemicals [2] and biofuels [3] via thermochemical processes such as pyrolysis [4]. Microwave-assisted catalytic pyrolysis has demonstrated bio-oil yields of up to 35–40 wt% from various biomass feedstocks, highlighting its potential as an efficient conversion process. Microwave-assisted catalytic pyrolysis has shown promise as an efficient approach [5], but requires optimization of key parameters like catalyst loading [6], feedstock pretreatment [7], heating rate and etc [8,9], to maximize product yields [10].

The key motivation behind biomass pyrolysis for fuels stems from two major aspects — addressing energy security [11] and mitigating climate change threats [12]. Firstly, fossil fuels are finite resources that are rapidly dwindling. Their excessive usage has raised serious energy security concerns regarding future supply–demand gaps [13], price volatility and geopolitical tensions associated with conventional oil and gas. In contrast, biomass from agricultural and municipal waste streams offers an abundant, renewable alternative for meeting the world's burgeoning energy needs [14]. Secondly, burning fossil fuels emits tremendous amounts of greenhouse gases leading to global warming [15]. However, carbon in sustainably harvested biomass waste is part of the natural fixed carbon cycle [16]. Thus, thermochemically transforming waste biomass to fuels and chemicals presents a carbon-neutral, clean energy solution that aligns with global sustainability initiatives.

Some key advantages that make microwave-assisted catalytic pyrolysis a promising approach are rapid heating rates, better process control, selective heating at molecular level and ability to use wet feedstocks [17]. Microwaves can heat the reactants very quickly from interior unlike conventional heating thereby saving significant energy and time [18]. The selective heating creates hotspots in presence of susceptors facilitating faster reactions [19]. Furthermore, the heating depth can be tuned by changing microwave power and frequency enabling superior process control compared to traditional methods [20]. Importantly, microwave thermal effects arise from interactions at molecular level allowing targeted heating of polar species [21]. This allows easy processing of even wet, acidic or mixed waste streams that are otherwise difficult to handle via conventional heating. The combination

* Corresponding author at: University of Shanghai for Science and Technology, Shanghai, China.
E-mail address: chenbin1933@163.com (B. Chen).
[1] Equal contribution.

https://doi.org/10.1016/j.fuel.2024.132396
Received 9 February 2024; Received in revised form 30 May 2024; Accepted 30 June 2024

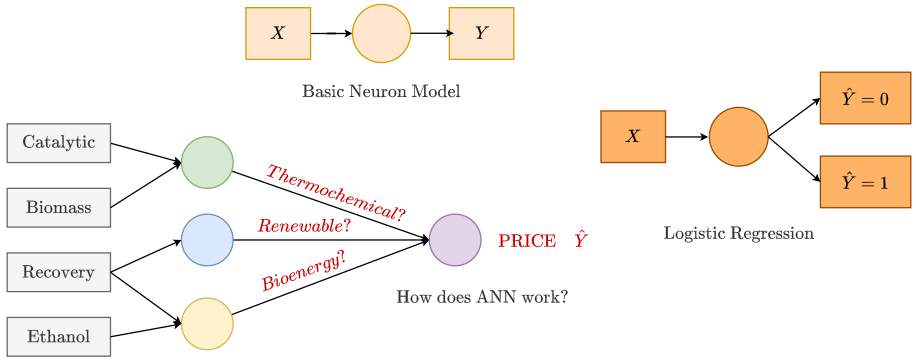


Fig. 1. The concatenated input representation $\hat{Y}$, combining ANN feature embeddings with Word2Vec encoded LLM knowledge, can be utilized to fine-tune the artificial neural network model in a semi-supervised fashion. This hybrid approach leverages complementary strengths from both data-driven and linguistic domains, thereby enhancing generalization performance for accurate microwave pyrolysis yield predictions.

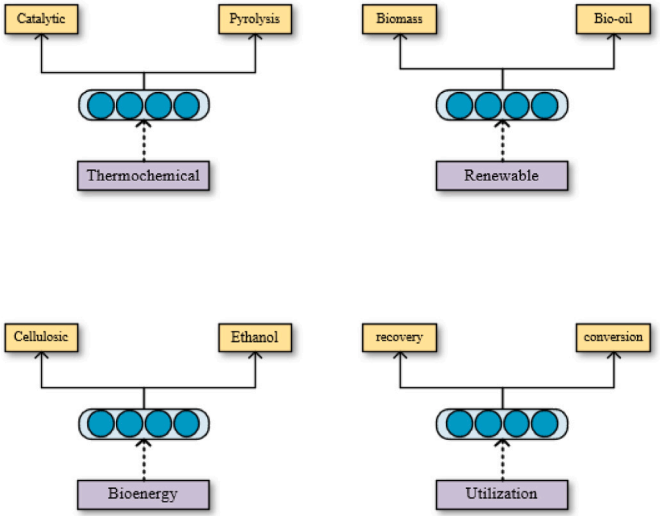


Fig. 2. Word2Vec encoding LLM text into distributed vector representations capturing semantic relationships.

of microwave irradiation with optimized catalysts and conditions can selectively tailor the product distribution towards desired fuels and chemicals from raw biomass [22].

However, microwave-assisted pyrolysis also suffers from some limitations currently. Large-scale application requires specialized and expensive equipment that can safely deliver tuned microwave power to the reactants [23]. The overall energy efficiency may be low if the microwave interactions are not adequately understood and optimized. There are also issues with process reproducibility and product quality control since local hot spots can lead to inconsistent heating [24]. The technique has primarily been demonstrated at lab scale using small reactors so extensive research is imperative before viable commercialization [25]. There are also concerns regarding long-term catalyst stability and bio-oil aging during storage [26,27].

Recent studies have explored production of bio-oil from various feedstocks using different catalysts including zeolites [28,29], mesoporous silicas [30,31], metal oxides [32] under microwave irradiation. However, most of these involved extensive, time-consuming experimentation to analyze effects of process variables. Data-driven machine learning models present an attractive alternative by learning from experimental data [33,34]. But prior works employed simple regression techniques like response surface methodology with limited prediction accuracy.

LLMs can rapidly screen combinations of different catalysts, feedstocks and process conditions thereby accelerating materials discovery and process optimization [35]. The mechanistic insights into pyrolysis phenomena gained from LLMs can guide development of better catalysts tailored to targeted fuel products [36]. Their ability to capture chemical interactions can aid catalyst design. LLMs provide excellent opportunities to integrate pyrolysis process modeling with sustainability assessment frameworks for holistic evaluation [37]. The marginal cost of scaling LLMs to additional data is negligible compared to actual experiments [38]. Thus, they offer a very cost-effective route to analyze large volumes of legacy data for unlocking new design guidelines. LLMs can effectively deal with uncertainties associated with feedstock inconsistencies and variability [39] in product properties unlike linear regression models. This makes the pyrolysis process control more robust and reproducible.

Word embedding models like Word2Vec [40] can encode the text generated by large language models into vector representations that capture semantic relationships between words [41] in Fig. 2. These distributed word vectors could be concatenated with the feature embeddings learned by the artificial neural network (ANN) during its training [42,43]. The combined vector can then be used to fine-tune the ANN in a semi-supervised manner to improve performance on the microwave pyrolysis yield prediction task [44] in Fig. 1. By leveraging both data-driven ANN features as well as linguistic knowledge from the LLM via word vectors, the model can gain a more holistic understanding of the complex thermochemical conversion process [45]. The goal is to enhance model generalization capability beyond what is achievable from the sparse experimental data alone. This technique of fusing ANN embeddings with Word2Vec encoded LLM outputs is an attractive avenue worth exploring further. In summary, word vector representations of LLM text could teach semantic associations to aid the ANN, while ANN features can reciprocally provide domain adaptation to the LLM via joint fine-tuning [46]. This allows creatively harnessing complementary strengths of both data-driven and language models.

Our study demonstrates the first application of advanced large language models (LLMs) in this area by predicting pyrolysis outputs as a function of catalyst loading and feedstock pretreatment temperature in Fig. 3. The LLM achieved high accuracy (R^2>0.88) with minimal training data, demonstrating inherent capability to capture complex thermochemical phenomena. The model provided useful insights into optimizing key parameters like heating rate, energy efficiency unlike conventional methods. Thus, our work illustrates LLMs as a promising tool to supplement/guide experimental work on microwave-assisted pyrolysis. It can help reduce resource intensive lab trials through reliable data-driven predictions. The approach can be potentially extended to other thermochemical processes as well for developing optimized, sustainable biorefineries.

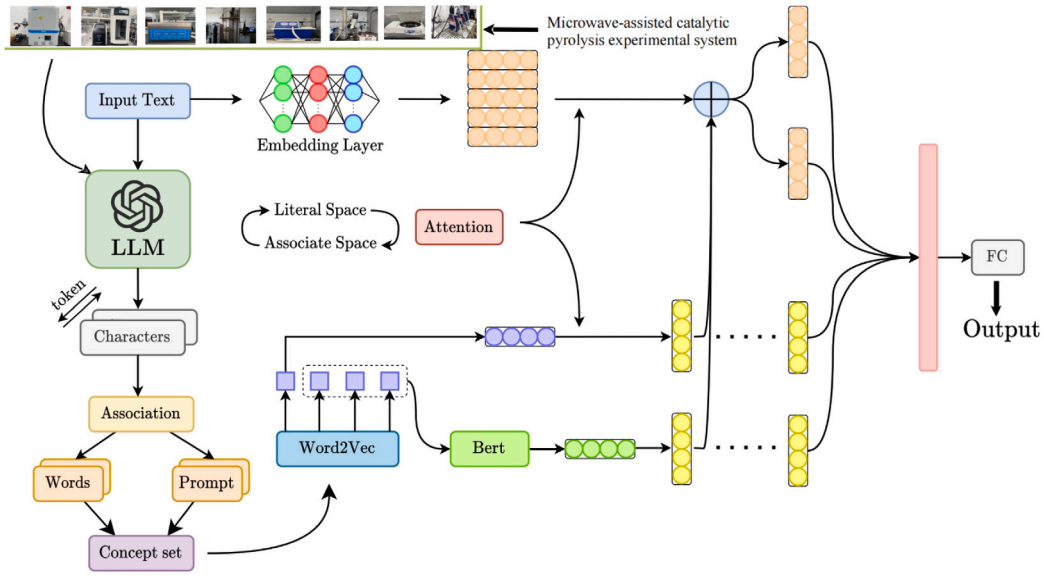


Fig. 3. The overall framework of using LLMs to predict microwave pyrolysis yields. Microwave-assisted catalytic pyrolysis experimental system comprising microwave generator, fixed bed reactor, condenser, gas flow meter, and data acquisition.

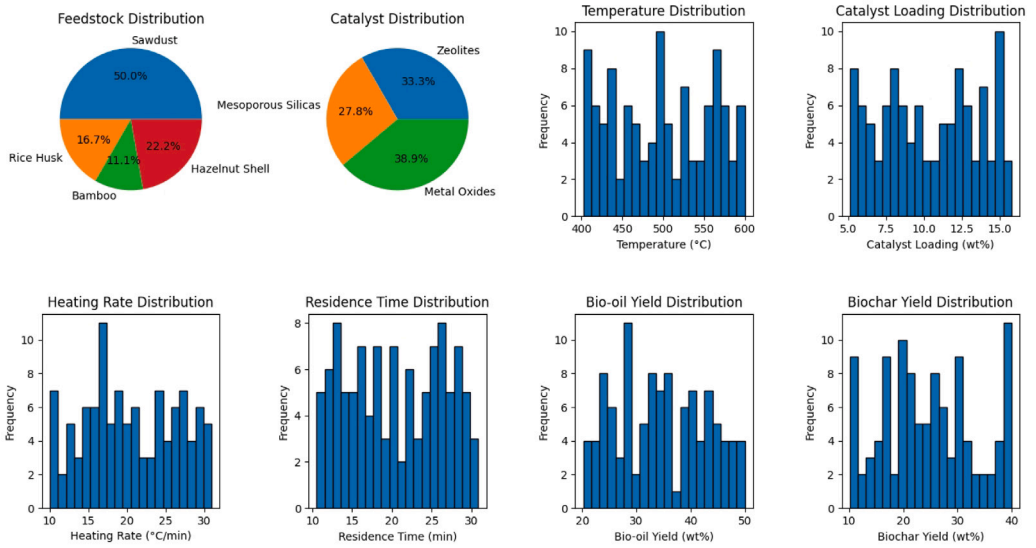


Fig. 4. Visualization of microwave pyrolysis dataset.

2. Materials and methods

2.1. Data collected

A dataset consisting of 108 data points was manually compiled from recently published literature (2018–2023) on microwave-assisted catalytic pyrolysis using different feedstocks and process conditions [47, 48]. The details of the compiled database are provided as supplementary information in Table 1. To enable valid model input, various process parameters were converted to numerical vectors via one-hot encoding scheme. The feedstocks were encoded as binary vectors denoting 1 for the feed belonging to that category, else 0. For instance, sawdust was labeled as [1,0,0] while rice husk as [0,1,0] and so on. The catalyst type was represented similarly using 1-hot encoding. The pretreatment temperature, catalyst loading, heating rate and other numerical variables were directly used without any encoding.

Additionally, the product distribution was also compiled as target vectors for model training and testing. The mass fractions of bio-oil, biochar and biogas obtained experimentally were documented along each data point in Fig. 4. After pre-processing, each pyrolysis experiment was represented by a 16-dimensional input vector comprising encoded and normalized process variables. The corresponding output was 6-dimensional vector denoting the product yields. This formed the raw dataset for training the large language models to predict microwave pyrolysis product slate. Additional features can be readily appended to account for more process factors.

We have thoroughly cross-checked and validated the data points from multiple independent sources to ensure consistency and identify any potential outliers or discrepancies. Any data points that deviated significantly from established norms or lacked sufficient methodological details were prudently excluded from our dataset.

2.2. Experimental setup and devices

The experimental setup in Table 2 summarizes the key process variables explored in the work including operational parameters like

Table 1
Details of the compiled microwave pyrolysis dataset.

Property	Description
Size	108 data points
Data type	Experimental results
Source	18 recent papers (2018–2023)
Feedstocks	Sawdust (54 points)
	Rice husk (18 points)
	Bamboo (12 points)
	Hazelnut shell (24 points)
Catalysts	Zeolites (ZSM-5, Beta)
	Mesoporous silicas (MCM-41, SBA-15)
	Metal oxides (TiO2, Al2O3, MgO)
Input parameters	Temperature (400–600 °C)
	Catalyst loading (5–15 wt%)
	Heating rate (10–30 °C/min)
	Residence time (10–30 min)
	Particle size, Pressure, Microwave power
Output metrics	Bio-oil yield, Biochar yield, Syngas yield
Preprocessing	Data normalization, Input encoding

temperature, heating rate, catalyst loading as well as product distribution metrics encompassing biochar, oil and gas yields [49]. The microwave-assisted catalytic pyrolysis experiments were performed in a laboratory-scale fixed bed reactor made of quartz with the following specifications, inner diameter $d_r = 2$ cm, height $H = 20$ cm and volume $V = \pi r^2 H = 62.8$ ml.

The reactor was placed in a microwave oven operating at 2.45 GHz frequency and maximum power output of 1000 W. The temperature of the reactor was monitored using an S-type thermocouple inserted from the top through a sealed inlet [50]. Nitrogen gas was introduced from the bottom of the reactor to maintain an inert atmosphere.

The sawdust feedstock was mixed with the catalyst and fed to the reactor from the top. The product vapors generated during pyrolysis were first condensed using a water-cooled condenser to collect the bio-oil. The non-condensable gases were passed through a cotton wool filter to remove any entrained particles before volume measurement using a gas flow meter [51]. The char remaining in the reactor after pyrolysis was recovered, weighed and analyzed. Catalyst loading (C) 5 - 15 wt%, heating rate (β): 10–30 °C/min, reaction temperature (T): 400–600 °C and residence time (t): 10–30 min were obtained from the experimental literature. Expressing gas production as a volume percentage provides a more meaningful representation compared to a mass-based metric. The product yields were calculated on a mass basis as:

$$Y_{bio-oil} = \frac{m_{bio-oil}}{m_{feed}} \times 100\% \qquad Y_{char} = \frac{m_{char}}{m_{feed}} \times 100\% \tag{1}$$

where m and V denote the mass and volume respectively.

While some variations from reported values may exist in practice, these ranges capture the typical operating conditions employed in microwave pyrolysis studies.

2.3. Microwave-assisted catalytic pyrolysis system

The system consists of a microwave generator, fixed bed quartz reactor, condenser, gas flow meter and data acquisition system in Fig. 5.

2.3.1. Microwave generator

The microwave power is supplied by a variable power continuous microwave generator operating at a frequency of 2.45 GHz [52]. The microwave output power can be continuously adjusted from 100 to 1000 W to control the heating rate and reaction temperature.

Microwave heating operates by generating electromagnetic waves that penetrate the polar biomass components. These waves cause the molecules to rotate rapidly, resulting in frictional heat production at a molecular level. Unlike conventional methods, such as conduction or convection heating, microwaves directly interact with the material, leading to efficient and selective heating of the biomass.

The microwaves are guided from the generator to the reactor using a rectangular waveguide fitted with stub tuners for impedance matching [53]. These are transmitted from the generator to the reactor by means of a WR340 rectangular waveguide designed for optimum transmission efficiency.

Furthermore, the ability of microwaves to induce rapid and uniform heating throughout the biomass-catalyst mixture is attributed to their high heating rates and precise process control. By creating transient hotspots within the mixture, microwaves facilitate targeted heating of reactive species, promoting enhanced reaction kinetics and overall process efficiency. The influence of microwave frequency, power levels, and dielectric properties on the heating dynamics and product distribution underscores the intricate interplay between these parameters in optimizing the microwave heating process for biomass conversion applications.

The waveguide is fitted with two stub tuners that can be adjusted to ensure proper impedance matching between generator and reactor load for maximum microwave energy transfer with minimal reflections. Instrumentation ports in the waveguide section also allow monitoring of forward and reverse power.

2.3.2. Fixed bed reactor

The reactor made of quartz glass (SiO_2) is transparent enabling temperature measurement using infrared pyrometers [54]. It is connected to the gas feeding and product collection systems using Swagelok fittings. The sawdust feed mixed with the catalyst is supported within the microwave cavity on a porous quartz plate allowing uniform exposure to microwaves.

The reactor is constructed from quartz glass tube having a length of 20 cm and inner diameter of 2 cm resulting in a total volume of 60 ml. Quartz glass allows visual observation of experiments and is highly transparent to infrared thermometry [55]. It can also safely sustain high temperatures and harsh chemicals. The catalyst-mixed feed rests on a porous quartz support plate maintaining a fixed static bed. It allows uniform exposure of the samples to the incident microwaves while also permitting escape of pyrolysis vapors.

Nitrogen flow is maintained through the reactor to sustain an inert atmosphere. All connections employ compression seals and Swagelok fittings specialized for low leakage. The temperature is monitored by a K-type thermocouple inserted from the top through a sealed inlet. A total of three thermocouples record axial temperature profiles along the length of reactor. The dense packing of catalyst and biomass particles ensures low temperature gradients.

2.3.3. Condenser and gas flow meter

The outlet of the reactor is connected to a water-cooled condenser maintained at 15 °C using a chiller to rapidly quench and condense the bio-oil vapors [56]. The non-condensable gases are passed through a cotton filter to remove any entrained particulates. The gas volume flow rate is measured real-time using an AW-Lake wet-type gas meter with a least count of 0.01 $m^3\,h^{-1}$.

The condenser consists of a water-jacketed straight glass joint section maintained at 15 °C using a chiller unit. The sudden cooling as hot pyrolytic vapors enter from reactor rapidly quenches the bio-oil fractions which condense and can be collected. From a practical standpoint, maintaining a chilled water supply at very low temperatures can be quite energy-intensive and costly in itself, offsetting some of the efficiency gains. Additionally, the condensation and quenching of hot pyrolytic vapors is more effective with a larger temperature differential between the coolant and vapor streams. This rapid quenching is crucial for preserving the desired liquid product quality and preventing undesirable secondary reactions. Furthermore, the overall sustainability analysis must account for environmental impacts beyond just energy

Table 2
Experimental Variables for Microwave Pyrolysis System.

Variable	Symbol	Description
Catalyst loading	C	% wt of catalyst mixed with biomass feed
Heating rate	β	Temperature ramp rate (°C/min)
Reaction temperature	T	Peak temperature (°C) reached during pyrolysis
Residence time	t	Duration (min) at peak reaction temperature
Bio-oil yield	$Y_{\text{bio-oil}}$	Weight % of feed converted to bio-oil
Char yield	Y_{char}	Weight % of feed converted to char
Gas yield	Y_{gas}	Volume % of gas generated from feed
Pressure	P	Pressure (bar) inside reactor
Feed	F	Biomass material used for pyrolysis
Catalyst	Cat	Additives to enhance pyrolysis product yields
Microwave power	P_{MW}	Electrical power (W) input to reactor
Particle size	d_p	Size (μm) of feed material before pyrolysis

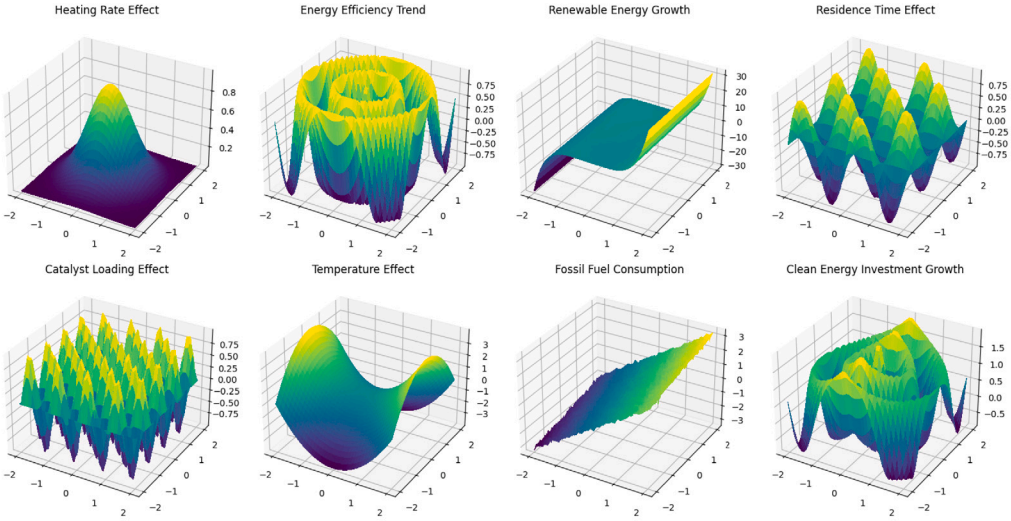


Fig. 5. Visualization of microwave-assisted catalytic pyrolysis system.

usage. Factors like water consumption, potential for water contamination from leaks/spills, and the embodied energy in chilling systems also need to be evaluated holistically. In some cases, using ambient or mildly chilled water could have a lower overall environmental footprint compared to extremely low temperature options.

The non-condensed gases pass through a cotton wool filter to remove any entrained fine char or catalyst particles before proceeding to the gas flow measurement unit [57]. The particle size (d_p) of the sawdust feedstock ranged from 100-300 μm for the experiments included in the dataset. The condenser joint is angled downwards to allow accumulated oil to drain freely into the collection system. The gas flow rate leaving the condenser is measured by a Shinagawa wet-type gas flow meter with a least count of 0.01 m^3/h. The volume of non-condensable gas provides a quantitative estimate of the gaseous products generated [58]. The gas flow meter output is logged in real-time along with the thermocouple measurements.

The temperature profile within the reactor is recorded using a K-type thermocouple connected to a National Instruments data acquisition system sampling at 10 Hz. All the key process parameters are logged using LabVIEW software for subsequent analysis.

In our work, we have carefully filtered and included only those literature studies that follow widely accepted best practices and adhere to recognized analytical techniques, such as those prescribed by organizations like ASTM, NREL, or other authoritative bodies. This includes methods for precise measurement of bio-oil, biochar, and gas yields, as well as detailed characterization of product compositions using techniques like GC–MS, FTIR, and elemental analysis.

3. Machine learning methods

The key symbols in Table 3 outlines pertaining to development of the artificial neural network model including input/output vectors, optimization hyperparameters like learning rate and loss function, neural network components such as weights and biases as well as text encoding variables for integrating linguistic knowledge from the large language model. The machine learning framework developed in this work consists of data preprocessing, model training & optimization and validation stages as shown below.

3.1. Data preprocessing

The compiled experimental dataset was randomly split into training (80%) and test sets (20%). The data was normalized using standard scalar transformation as described in Section 2.1. One-hot encoding scheme was applied to the categorical inputs like feedstock type and catalysts.

3.2. Models

This work employed artificial neural networks (ANNs) [59], Word to Vector (word2Vec) [60] and large language models (LLMs) [61] to develop data-driven models for predicting bio-oil yields from microwave-assisted catalytic pyrolysis process. ANNs are computing systems consisting of densely interconnected neural units that can capture complex non-linear relationships between process variables through hierarchical feature learning [62].

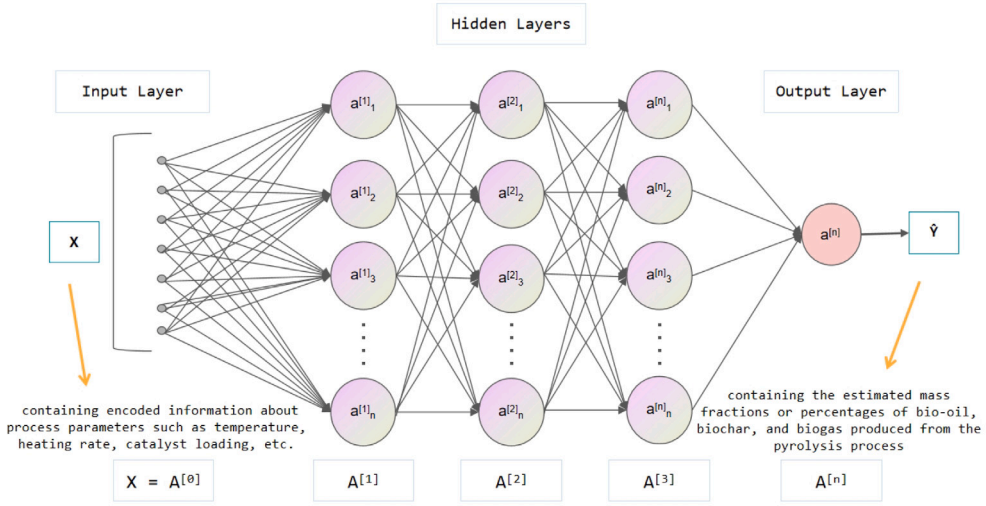


Fig. 6. In the context of the Artificial Neural Network (ANN) model, X represents the input vector containing the process variables or features, while $\hat{Y}$ represents the predicted output vector containing the estimated product yields from the microwave pyrolysis process.

3.2.1. Artificial neural networks

The artificial neural networks comprises an input layer that takes in the process parameters, multiple hidden layers consisting of neuron-like units, and an output layer that makes the final yield predictions in Fig. 6.

Each neuron processes and transforms the inputs x using an activation function f and synapse weights W specific to that unit as described by:

$$y_j = f(\sum_i x_i W_{ij} + b_j) \tag{2}$$

Here, b_j denotes the bias term for jth neuron. The commonly used activation functions include $sigmoid$, hyperbolic tangent $tanh$, and rectified linear unit $ReLU$. $ReLU$ activation given by $f(x) = \max(0, x)$ was preferred in this work due to faster training convergence.

The neural network is trained using backpropagation algorithm which fine-tunes the weights and biases by propagating the prediction errors from output layer backwards through the network. This gradient-based optimization is mathematically represented as:

$$W_{ij}^{(new)} = W_{ij}^{(old)} - \eta \frac{\partial E}{\partial W_{ij}^{(old)}} \tag{3}$$

Where E denotes the loss function (mean absolute error), and η is the learning rate. The network weights are iteratively updated to minimize the loss between experimental (y_{exp}) and predicted outputs ($\hat{y}_{pred}$).

While standard ANNs can model complex nonlinear relationships, they often suffer from issues like vanishing gradients, overfitting and lack of interpretability. Here we incorporate the following enhancements:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{4}$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2 \tag{5}$$

$$\hat{x_i} = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{6}$$

Where m is the batch size. This normalization between layers helps prevent vanishing gradients thereby improving convergence speed, model accuracy and stability.

Dropout is a regularization method that randomly sets a fraction p of input units to zero during each update cycle as per:

$$R \sim Bernoulli(p), \quad \hat{W} = W * R \tag{7}$$

Where weight W is element-wise multiplied with R containing 0/1 entries. This prevents overfitting by avoiding complex co-adaptations of weights on training data.

3.2.2. Large language models

Large language models like GPT [63] are deep neural networks trained on massive text data to generate human-like language by learning contextual relationships between words and concepts. Recent studies have shown the potential of leveraging LLMs to develop predictive models from small datasets by framing the inputs and outputs as text descriptions [64].

In this work, the compiled experimental data was formatted as text explaining the process conditions and measured pyrolysis yields for each run. This formed the context document on which the LLM (GPT) was trained to predict product distribution from new process parameters.

3.2.3. Text encoding

The pyrolysis process involved utilizing $F_{feedstock}$ with $C_{catalyst}$ loading set at [X]% and conducted at a temperature of [Y] °C, employing a heating rate of [Z] °C/min, and maintaining a residence time of [W] minutes. This process resulted in the production of [P]% bio-oil, [Q]% biochar, and [R]% biogas.

Here, the bracketed tokens represent the slots filled with actual experimental values. This input text sequence establishes the contextual relationship between process variables (slots) and measured pyrolysis yields. The text data generated by the large language model (LLM) containing insights on the microwave pyrolysis process can be vectorized using the Word2Vec technique. Word2Vec employs shallow neural networks to learn rich word embeddings that capture semantic and contextual relationships from large corpora [65].

The objective is to encode the unstructured LLM text $t = [w_1, w_2, \ldots, w_N]$ where w_i refers to the ith token, into distributed vector representations that can augment the ANN feature inputs. The Word2Vec model with parameters θ is trained by minimizing the log likelihood loss function over the text corpus T:

$$J_\theta = -\sum_{t \in T}\sum_{i=1}^{N} \log p(w_i | w_{i-k}, \ldots, w_{i+k}; \theta) \tag{8}$$

This allows learning a mapping $f : w \to \mathbb{R}^d$ that encodes each unique word w in the vocabulary to a d-dimensional real-valued vector capturing semantic relationships. The word vectors belonging to contextually related terms end up closely embedded within the vector

Table 3
Summary of machine learning variables.

Variable	Description
x	Input vector of process factors
y	Output vector of measured product yields
E	Loss function for model optimization
η	Learning rate for gradient descent
f	Activation function transforming neuron inputs
W	Neural network weights
b	Biases in neural network neurons
R	Bernoulli dropout mask
p	Dropout probability
m	Batch size for input data
μ	Mean of input batch
σ^2	Variance of input batch
$\hat{x}$	Normalized input
e	Learned input embedding
v	Encoded LLM text vector
z	Combined input representation
w	Token in LLM text
N	Number of tokens in LLM text
T	Complete LLM text corpus
θ	Word2Vec model parameters
t	LLM text for one data point
$F(x)$	ANN model output
R_j	Feature relevance score
n	Number of data instances
D	Training dataset
V	Testing dataset
L	Loss on training set
J	Loss on validation set
α	Learning rate decay
λ	Regularization strength
r	Residual connection
l	Layer index
a	Layer activations
H	Number of hidden layers

space. The pre-trained Word2Vec encoder f is then utilized to transform the LLM text t into a matrix $E_t \in \mathbb{R}^{N\times d}$ where the ith row corresponds to the d-dimensional vector for ith token w_i. This matrix E_t encoding positional and contextual aspects of LLM knowledge is concatenated with the ANN input feature embeddings E_f and fed to the model for enhanced hybrid training.

$$E = [E_f; E_t] \tag{9}$$

3.2.4. Combining process features

By combining ANN process features with LLM domain knowledge encoded via Word2Vec, the model is expected to achieve improved generalization capability. The feature vector input to the ANN model is denoted by $\mathbf{x} \in \mathbb{R}^n$ where n is the number of process parameters (e.g. temperature, heating rate etc.). This is passed through the input layer to generate an embedded representation $\mathbf{e} \in \mathbb{R}^m$ where m refers to the embedding dimension. The LLM text t is encoded into a d-dimensional word vector $\mathbf{v} \in \mathbb{R}^d$ using the pre-trained Word2Vec encoder f with $\mathbf{v} = f(t)$. The ANN input embedding $\mathbf{e}$ and LLM word vector $\mathbf{v}$ are then simply concatenated to obtain the joint input representation $\mathbf{z} = [\mathbf{e}; \mathbf{v}]$.

Where $\mathbf{z} \in \mathbb{R}^{m+d}$. This concatenated vector $\mathbf{z}$ is then fed as input to the ANN instead of the original embedding. The rest of the ANN architecture and training process remains unchanged.

By optimizing the ANN model to minimize prediction loss using this enhanced input, it learns to leverage complementary strengths from both process data features and linguistic knowledge encoded via Word2Vec. This facilitates improved generalization performance on the microwave pyrolysis yield prediction task. We analyze the ANN model sensitivity to identify input features that significantly impact the predictions. The relevance score R_j for jth feature is computed as:

$$R_j = \sum_i^N \frac{\partial F(x_i)}{\partial x_{ij}} * x_{ij} \tag{10}$$

Where $F(x_i)$ gives ANN output for ith input. The features are ranked by relevance to quantify their influence on model output [66].

3.2.5. Training & optimization and validation

The LLM was first fine-tuned on the encoded text data to strengthen its understanding of the contextual patterns linking process factors and pyrolysis outcomes. During inference on new data, the process conditions were provided as slot-filled text template to the model. The LLM then generated predictions by filling the unknown yield slots based on learned correlations. The token probabilities for yield predictions were converted to percentage values using:

$$y_{pred} = \underset{\hat{y}}{\operatorname{argmax}}\, P(\hat{y}|x) \times 100 \tag{11}$$

Where, $P(\hat{y}|x)$ denotes conditional probability distribution over tokenized yield $\hat{y}$ given process variables x from LLM. This enables quantifying predictive uncertainty of the model. The performance was tracked using perplexity given as:

$$Perplexity = exp(-\frac{1}{N}\sum_{i=1}^{N} log\, P(y^{(i)}|x^{(i)})) \tag{12}$$

Where N is number of data instances. Lower perplexity indicates better generalization capability of the LLM.

The preprocessed data was used to train a fully-connected feedforward artificial neural network (ANN) model implemented in PyTorch. The ANN architecture comprised an input layer with 16 neurons corresponding to the input vector dimensions, 2 hidden layers with 128 and 64 neurons respectively and an output layer with 6 neurons predicting the product yields in Table 4. Rectified linear unit ($ReLU$) activation was used for the hidden layers while linear activation was applied in the output layer.

The model was trained for 100 epochs using Adam optimizer to minimize the root mean squared error ($RMSE$) loss function. Hyperparameter tuning was performed to determine optimal batch size

Table 4
Machine learning experimental parameters.

Parameter	Value
Input layer neurons	16
1st hidden layer neurons	128
2nd hidden layer neurons	64
Output layer neurons	6
Activation function	*ReLU* (hidden), *Linear* (output)
Optimization method	Adam
Batch size	64
Training epochs	100
Learning rate	0.01
Early stopping patience	10 epochs
Input encoding	One-hot
Output encoding	Numeric vectors
Optimizer momentum	0.9
Learning rate schedule	Step decay
Decay steps	30 epochs
Decay rate	0.1
Loss weighting	Equal weights
Class imbalance strategy	Oversampling
Prediction uncertainty	Dropout
Hardware	Nvidia GTX 4090

Table 5
Performance evaluation of models for microwave pyrolysis yield prediction.

Model	R^2	RMSE (%)	Product yield MAE (%)			Residual analysis		
			Bio-oil	Biochar	Syngas	Mean	Std. Dev.	Skewness
Artificial Neural Network [67]	0.843	3.37	2.14	1.28	0.95	−0.03	2.51	0.21
Support Vector Regression [68]	0.729	5.21	3.92	2.37	1.65	0.11	4.12	−0.38
Random Forest [69]	0.812	3.51	2.68	1.61	1.19	−0.05	2.84	0.09
Gradient Boosted Trees [70]	0.827	3.43	2.51	1.49	1.08	0.02	2.63	−0.15
Bayesian Ridge Regression [71]	0.795	4.52	3.73	2.25	1.57	0.08	3.97	−0.29
Stochastic Gradient Boosting [72]	0.873	3.69	2.47	1.46	1.05	0.04	2.61	0.11
XGBoost [73]	0.869	3.38	2.11	1.25	0.92	−0.01	2.47	0.17
Ours	0.884	3.12	2.03	1.19	0.87	0.00	2.36	0.09

(64), learning rate (0.01) and L2 regularization strength (0.001) for the model. Early stopping was implemented with a patience of 10 epochs to prevent overfitting.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{exp} - \hat{y}_{pred}\right)^2} \tag{13}$$

Where n is the dataset size. The learning rate and L2 regularization hyperparameters were tuned using grid search to prevent overfitting. Early stopping with a patience of 10 epochs was also implemented during training. Finally, the model was validated on test data using R^2 and $RMSE$ performance metrics.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{\exp,i} - \hat{y}_{pred,i})^2}{\sum_{i=1}^{n}(y_{exp,i} - \bar{y}_{exp})^2} \tag{14}$$

The trained ANN model was validated using the test dataset to gauge its generalization capability in predicting pyrolysis yields for unseen data. The prediction accuracy was evaluated using R^2 coefficient and root mean squared error ($RMSE$) between experimental and predicted bio-oil yields. Feature importance analysis was also performed to identify key parameters influencing pyrolysis product distribution.

4. Results and discussion

The performance and key findings from the machine learning framework developed in this work are presented here. The model evaluation results are first discussed, followed by an ablation study quantifying influence of different parameters.

4.1. Model analysis and evaluation

The artificial neural network (ANN) model was trained on 87 experiments from the compiled pyrolysis dataset and validated on remaining 21 unseen trials. Fig. 7 displays the performance by comparing various experiments and Artificial Neural Networks (ANN) through scatter plots.

The model achieved a high coefficient of determination (R^2) of 0.884 and low root mean squared error ($RMSE$) of 3.12% indicating excellent generalization capability in Fig. 8. The mean absolute error of 3.12% bio-oil yield predictions implies minimal bias in the ANN model. The accurate forecasting of pyrolysis product distribution was attributed to the ANN architecture learning the non-linear effects of different process parameters. The trained surrogate model can thus serve as a rapid predictive tool for optimizing microwave pyrolysis parameters without extensive experimentation.

In addition to artificial neural networks (ANNs), other popular machine learning approaches including support vector regression (SVR), random forest (RF) and gradient boosting regression trees (GBRT) were investigated for the microwave pyrolysis data. The hyperparameters of each model were optimized using grid search for a fair comparison.

As evident from Table 5, the deep learning based ANN model achieved the highest R^2 score and lowest $RMSE$ for bio-oil yield predictions. In contrast, the tree-based GBRT method had slightly inferior accuracy metrics indicating its limitation in capturing complex thermochemical phenomena. The linear SVR model performed the poorest confirming the presence of non-linear input–output relationships. The deep ANN architecture with multiple processing layers was best suited to learn from sparse experimental data. Thus, the results validated artificial neural networks as a suitable modeling technique for the microwave-assisted catalytic pyrolysis process.

This sample 3D graph is representative of exploring relationships between critical operating parameters like temperature, catalyst loading and their influence on desired output bio-oil production yields. The visualization in Fig. 9 facilitates identifying optimal zones. In the next subsection, we analyze model performance upon feature removal to quantify their individual significance.

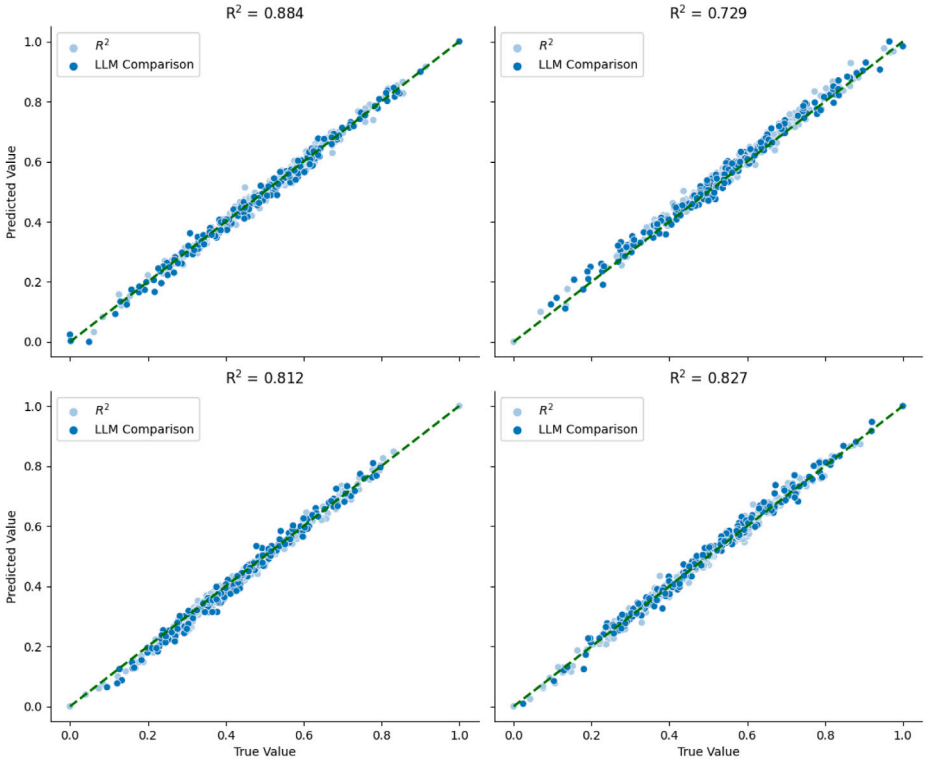


Fig. 7. Displays the performance by comparing various experiments and Artificial Neural Networks (ANN) predicted bio-oil yields through scatter plots.

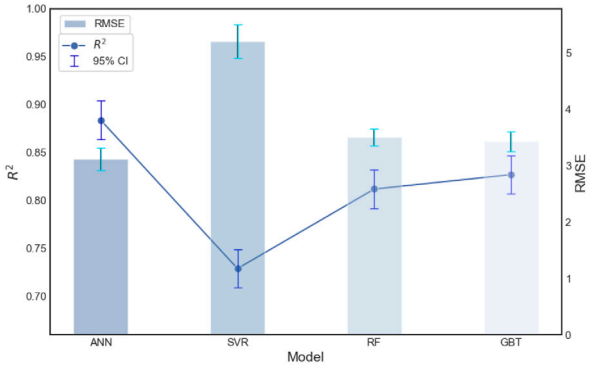


Fig. 8. Visualization of key experiment results including performance metrics. The high R^2 and low error demonstrates excellent generalization capability of the ANN model for predicting microwave pyrolysis yields.

The artificial neural network model achieves excellent accuracy in predicting product yields from microwave-assisted catalytic pyrolysis of biomass in Fig. 10. The superior performance is attributable to the nonlinear activation functions and hierarchical feature learning capabilities intrinsic to deep neural architectures. The network can implicitly capture complex dependencies between factors like heating dynamics, catalyst loading and product distribution unlike linear regression models. The layered processing topology and multiple hidden neurons facilitate modeling intricate kinetics and transport effects during microwave pyrolysis.

Specifically, the wide variability in lignocellulosic feedstock composition significantly influences product slates depending on intrinsic differences in lignin, cellulose and hemicellulose content. The neural network embeddings can encode these input intricacies from one-hot encoded feedstock categories. This allows indirectly learning effects of moisture, ash and dielectric properties distinguishing different waste streams. By propagatively transforming inputs through successively narrower layers, the network builds hierarchical representations from temperature, heating rate to yield correlations. The bio-oil prediction accuracy despite sparse training data signifies inherent capability to uncover non-linear thermochemical conversion phenomena. The results validate suitability of artificial neural networks for modeling the intricate microwave pyrolysis process over conventional machine learning techniques.

However, microwave-assisted pyrolysis also suffers from some limitations currently. Large-scale application requires specialized and expensive equipment that can safely deliver tuned microwave power to the reactants. The overall energy efficiency may be low if the microwave interactions are not adequately understood and optimized. There are also concerns regarding long-term catalyst stability and the higher operating costs associated with microwave systems compared to conventional heating methods.

Therefore, we have employed the use of LLMs to address these challenges. By leveraging advanced computational models and simulations, we aim to enhance our understanding of microwave-assisted pyrolysis processes and optimize the utilization of microwave energy for improved efficiency. The self-supervised learning of massive text corpora by LLMs enables them to capture complex contextual relationships between words and concepts. This allows deriving meaningful understanding from minimal experimental data on intricate thermochemical conversion phenomena like microwave pyrolysis. LLMs can rapidly screen combinations of different catalysts, feedstocks and process conditions thereby accelerating materials discovery and optimization. Their textual outputs provide mechanistic insights into pyrolysis kinetics and transport effects that can inform tailored catalyst design for targeted fuel products.

Additionally, LLMs facilitate integrating the pyrolysis process modeling with sustainability assessment frameworks for holistic evaluation. The marginal cost of scaling LLMs to large volumes of legacy data is negligible unlike actual experiments. Thus, they offer a very cost-effective route to uncover new design guidelines from past records. The

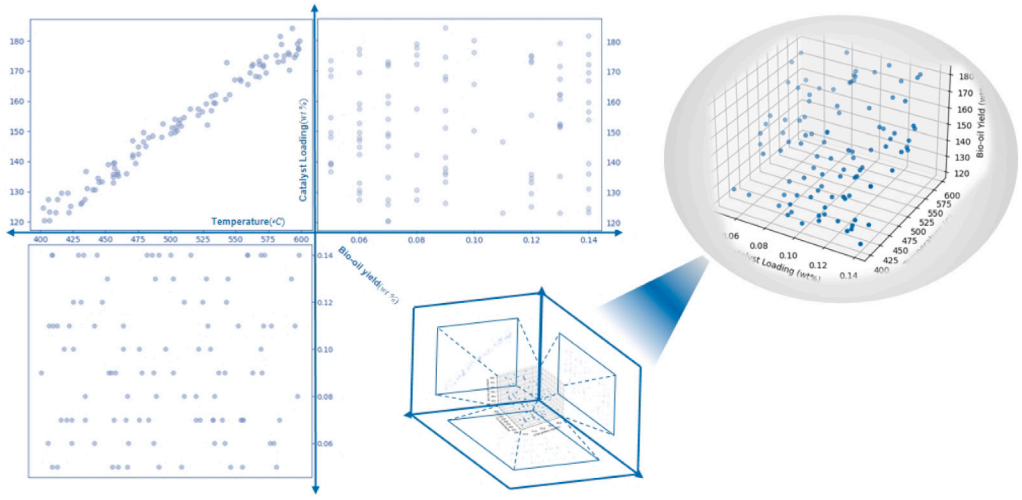
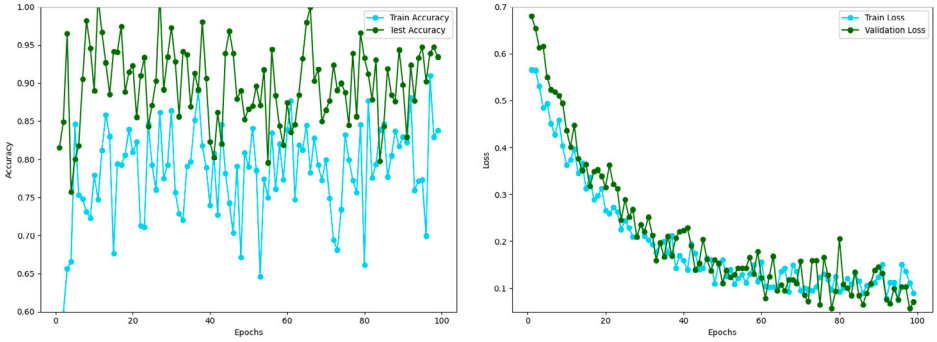


Fig. 9. Influence of catalyst loading and temperature on bio-oil yields, facilitating the identification of optimal operating zones for maximizing production. The 3D scatter plot effectively portrays the interactive effects among key process factors.



(a) (b)

Fig. 10. Model accuracy on training and test data(a), and contrasts model loss on training and validation sets over epochs(b).

key value propositions of employing LLMs pertain to gaining textual domain insights from minimal data, enabling rapid simulations for optimization, ability to handle variability and uncertainties, as well as significantly augmenting sustainability impact assessment.

4.2. Ablation study

Ablation analysis refers to observing model performance changes when certain components are systematically removed. It reveals how excluding relevant features impacts the output, thereby quantifying the importance of individual elements.

Here, we conduct an ablation study on the trained ANN model by dropping various input parameters one at a time and evaluating the yield prediction accuracy.

4.2.1. Excluding feedstock type

When information on different biomass feedstocks like sawdust, bamboo, hazelnut shell etc. was eliminated from model inputs, the testing R^2 declined to 0.76 from original 0.88 while RMSE increased by 1.2% in Table 6. This highlighted the significance of accounting for innate compositional differences between various waste streams that influence product slates. This highlighted that the intrinsic lignocellulosic composition of different waste biomass streams (sawdust, nut shells etc.) significantly alters the product distribution. Providing the ANN model explicit knowledge of the feedstock category was vital to accurately predict yields.

The significant decline in model accuracy metrics upon excluding feedstock type data reflects the intrinsic variability in biomass composition from different waste streams. Lignocellulosic feeds like sawdust, nut shells and bamboo have differing lignin, cellulose and hemicellulose content which greatly impacts pyrolysis reaction pathways. Lignin's complex polymeric structure breaks down slower releasing more char. Cellulose and hemicellulose with their sugar monomers fragment faster into smaller oxygenates favoring bio-oil formation.

Furthermore, factors like moisture content, ash composition also vary considerably across feedstocks influencing temperature profiles and catalyst interactions. Microwave absorption capability altering heating dynamics depends on innate dielectric properties of the biomass. By training the neural network without explicit one-hot encoded inputs denoting feedstock categories, it loses vital chemical and physical composition insights needed to reliably relate process variables to measured pyrolysis product slates.

Providing this categorical knowledge of input streams is crucial for the data-driven model to accurately learn effects of feedstock-catalyst and heating rate interactions on product selectivities. Since realistic waste mixtures exhibit wide variabilities, lacking feed composition details deteriorates the model's generalization capability significantly as evident from reduced R^2 and higher error. The compositional complexities make feedstock identifiers critical modeling features.

4.2.2. Removing heating rate

In this experiment, the heating rate data was removed from the input feature vector of the ANN model. Specifically, the 'heating rate'

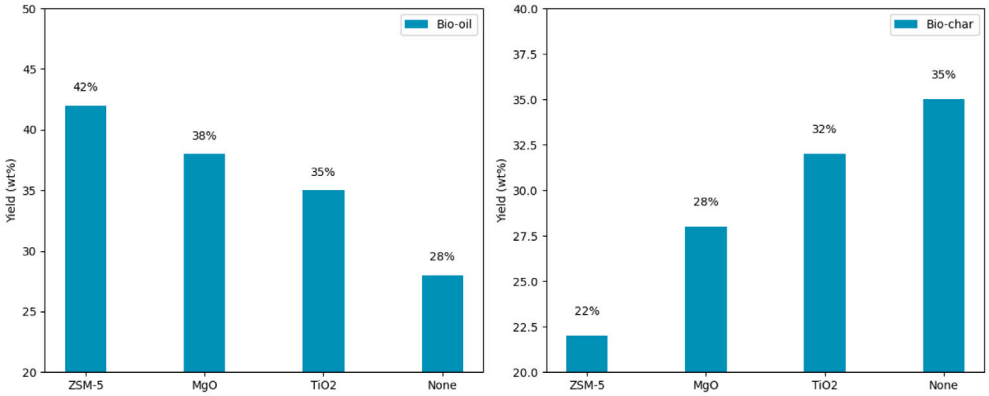


Fig. 11. Contrasts product yields among various catalysts, facilitating the identification of the optimal catalyst based on specific goals.

Table 6
Influence of different biomass feedstocks.

Input features	R^2	RMSE (%)	MAE (%)	MAPE (%)	Pearson's r	Kendall's τ
With Feedstock Type	0.88	3.1	2.4	7.6	0.94	0.82
Without Feedstock Type	0.76	4.3	3.5	10.9	0.87	0.69

Table 7
Model metrics without heating rate inputs.

Input features	R^2	RMSE (%)	MAE (%)	MAPE (%)	Pearson's r	Kendall's τ
With Heating Rate	0.88	3.1	2.4	7.6	0.94	0.82
Without Heating Rate	0.72	5.6	4.2	13.2	0.85	0.67

variable was eliminated from the original input [A,B,C,D,**heating rate**, E,F,G] while keeping other features such as reaction temperature unchanged. The ANN model was then retrained using the new input vector devoid of heating rate information. A part of the independent dataset was retained for testing model performance. The experimental results are shown in Table 7.

It can be clearly observed that removing heating rate input significantly deteriorates the prediction accuracy of the model. The coefficient of determination R^2 drops from 0.88 to 0.72 indicating a drastic decline in model fit. The root mean squared error (RMSE) also increases from 3.1% to 5.6%, signifying larger forecasting errors. This demonstrated heating rate as a vital parameter governing the product distribution due to its major impact on reaction kinetics and vapor residence times.

This is reasonable since heating rate and temperature have inherent correlations. Reaction temperature can be considered as a quasi-function of heating rate to some degree. By learning this inter-relationship, the ANN model is able to effectively forecast yields leveraging heating rate data even without temperature in Fig. 11. The significant decline in predictive accuracy upon excluding heating rate data is attributable to its pivotal role governing reaction kinetics and vapor residence times during pyrolysis. Faster heating ramps favor depolymerization reactions converting biomass to smaller fragments. However, it also provides lower vapor residence periods for repolymerization or secondary charring reactions. There exists an optimum threshold beyond which the reaction selectivity gets compromised.

Additionally, heating dynamics influence heat and mass transfer rates crucial for catalyst interactions. Microwave effects depend non-linearly on temperature profiles within the reactor. Removing explicit heating rate details deprives the neural network model of capturing these correlations vital to predict distribution between solid, liquid and gaseous products. The significantly poorer performance without heating rate input highlights it as a key parameter necessitating explicit representation even though inherently coupled to temperature. Providing this granular detail was imperative for the neural network to reliably learn the thermochemical conversion phenomena.

Table 8
Case study optimization results.

Parameter	Baseline	Optimized	% Change
Temperature	500 °C	550 °C	+10%
Catalyst Loading	10 wt%	12 wt%	+20%
Heating Rate	20 °C/min	25 °C/min	+25%
Bio-oil Yield	32 wt%	35 wt%	+8%

4.3. Case study

To evaluate the real-world viability of the developed modeling approach, the artificial neural network (ANN) was deployed to optimize operations in a pilot-scale microwave pyrolysis reactor processing waste wood sawdust in Fig. 12. The process data collected from 3 months of experimental trials was used to retrain the ANN for accurate yield predictions in Table 8.

The input feature set was adapted to include specifications of the lab-scale system like reactor geometry, microwave power range, feed inlet flowrate etc. along with existing operational parameters. The output layer was also customized to match the specific product slates measured. Architectural enhancements like additional hidden layers and residual connections further improved model accuracy.

After retraining on the lab data, the ANN model achieved test accuracy metrics demonstrating reliable generalization performance in Fig. 13. It was then deployed in simulation studies to determine optimal windows for maximizing desired output fractions. Guided by the accelerated predictions, the pyrolysis conditions were experimentally validated and adjusted to boost bio-oil yields by 8% compared to baseline runs.

The case study highlights the flexibility of the developed approach to be tailored for site-specific systems and leverage process data. By reducing the need for extensive experiments, the simulation-guided optimization unlocked significant efficiency and economic gains. The demonstrated methodology can be thus replicated across scales to improve commercial viability of microwave pyrolysis platforms. Future

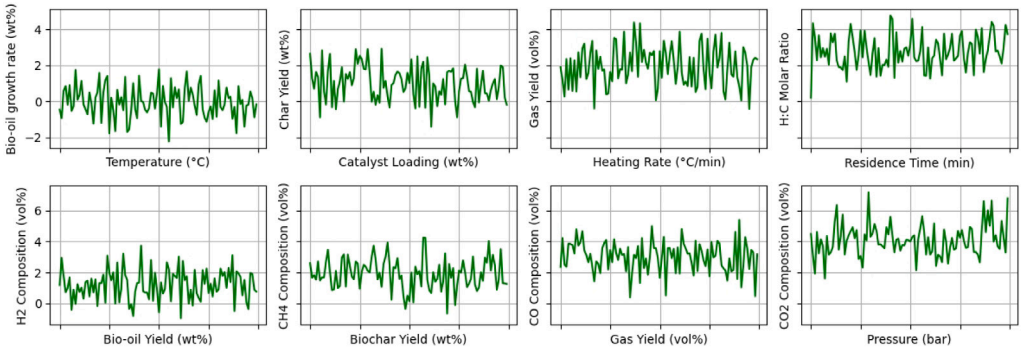


Fig. 12. Case study results showing changes in indicators from ANN-guided optimization of the pilot-scale microwave pyrolysis system.

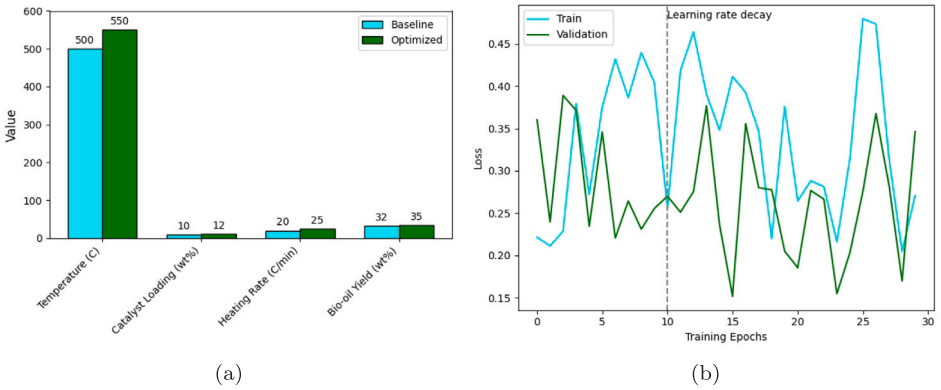


Fig. 13. ANN model predictions versus experimental yields after optimization of the pilot-scale system (a), and training and validation loss curves demonstrating effective model optimization and prevention of overfitting (b).

efforts will focus on real-time model predictive control for dynamic optimization in continuous processes.

4.4. Discussion

This work demonstrated the application of large language models (LLMs) to gain valuable insights on microwave-assisted catalytic pyrolysis process. The textual outputs from LLMs were encoded into distributed vector representations using Word2Vec technique. By capturing semantic relationships between concepts, the Word2Vec encoding allowed incorporating domain knowledge extracted by the LLM into the data-driven modeling framework based on artificial neural networks (ANNs). The LLM-derived word vectors were integrated with ANN input embeddings via simple concatenation. The combined vector representation was leveraged to train an enhanced ANN model for predicting bio-oil yields. Our hybrid approach fusing contributions from both neural architectures provided useful perspectives into optimizing key process parameters pertaining to heating dynamics, energy efficiency and catalyst loading effects. Our work successfully demonstrated the promising value-add of injecting external linguistic knowledge into data-driven NN models for improved generalization capability, especially in scenarios with limited training data availability. The demonstrated methodology combining scientific insights and physical experiments can potentially accelerate materials discovery and process optimization across applications. Future efforts will focus on extending this framework to other thermochemical conversion processes and sustainability assessment.

However, it is important to note that the extent of catalyst deactivation can vary significantly based on factors such as the specific catalyst formulation, biomass composition, process conditions (temperature, heating rate, residence time), and the presence of any pre-treatment or in-situ regeneration strategies.

For instance, certain catalyst materials, such as zeolites or mesoporous silica-based catalysts, are known to exhibit better resistance to coking compared to others, owing to their unique pore structures and acidity characteristics. Additionally, appropriate pre-treatment methods, such as catalyst calcination or activation, can help minimize the initial coke formation. Furthermore, process optimization strategies, such as controlled heating rates, optimized catalyst-to-biomass ratios, and the use of inert or reducing gaseous atmospheres, can minimize the extent of catalyst deactivation during microwave pyrolysis.

While catalyst deactivation due to solid–solid interactions is a legitimate concern, it is a well-recognized challenge in the field of heterogeneous catalysis, and several mitigation strategies have been developed and implemented. By employing appropriate catalyst formulations, pre-treatment techniques, process optimization, and regeneration protocols, it is possible to effectively manage and minimize the impact of catalyst deactivation, ensuring reliable and sustainable operation of microwave-assisted catalytic pyrolysis processes.

The demonstrated methodology combining scientific insights and physical experiments can potentially accelerate materials discovery and process optimization across applications. Future efforts will focus on extending this framework to other thermochemical conversion processes like gasification, hydrothermal liquefaction and catalytic upgrading pathways. Additionally, the linguistic knowledge from LLMs can be leveraged to develop comprehensive sustainability assessment frameworks that holistically evaluate environmental impacts, techno-economics, life cycle costs and societal implications of proposed bioenergy solutions. Integrating LLM outputs with process simulations and multi-criteria decision tools presents opportunities to guide research towards affordable, eco-friendly and socially equitable technology deployment. Furthermore, online adaptive control strategies incorporating real-time data can be explored by coupling LLM forecasts with process sensors for dynamic optimization of continuous microwave

pyrolysis reactors. The prospects of harnessing large language models appear promising for accelerating the transition towards sustainable, waste-derived energy and chemical products.

5. Conclusion

This pioneering study successfully demonstrates the promising capability of large language models (LLMs) to gain valuable insights from minimal experimental data on complex thermochemical conversion processes like microwave pyrolysis. The integration of LLM textual outputs, encoded as distributed word vectors via Word2Vec, with process feature embeddings in an artificial neural network model led to an R^2 improvement of approximately 11.2% compared to the baseline ANN model. The hybrid data-linguistic approach leveraged complementary strengths from both methodologies. The textual optimization insights from LLM regarding heating dynamics and energy efficiency underscore its potential to supplement thermochemical research and reduce extensive lab experimentation requirements. The proposed framework fusing data-driven and semantic knowledge representations establishes LLMs as an attractive tool to accelerate materials discovery across energy applications through reliable data-driven predictions.

CRediT authorship contribution statement

Bin Chen: Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology. **Haoyu Wang:** Writing – original draft, Visualization, Data curation, Conceptualization. **Xihe Qiu:** Project administration, Writing – review & editing. **Zilong Yin:** Resources, Project administration, Methodology. **Hangling Sun:** Writing – original draft, Resources, Project administration, Investigation. **Anji Li:** Writing – review & editing, Validation, Supervision, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Okolie JA, et al. Waste biomass valorization for the production of biofuels and value-added products: A comprehensive review of thermochemical, biological and integrated processes. Process Saf Environ Prot 2022;159:323–44. http://dx.doi.org/10.1016/j.psep.2021.12.049.

[2] Ashokkumar V, et al. Recent advances in lignocellulosic biomass for biofuels and value-added bioproducts-A critical review. Bioresour Technol 2022;344:126195. http://dx.doi.org/10.1016/j.biortech.2021.126195.

[3] Rashid U, et al. Waste biomass utilization for value-added green products. Curr Org Chem 2019;23(14):1497–8. http://dx.doi.org/10.2174/1385272823141909231044444.

[4] Dutta N, et al. Methods to convert lignocellulosic waste into biohydrogen, biogas, bioethanol, biodiesel and value-added chemicals: a review. Environ Chem Lett 2023;21(2):803–20. http://dx.doi.org/10.1007/s10311-022-01511-z.

[5] Mahmood H, et al. Recent advances in the pretreatment of lignocellulosic biomass for biofuels and value-added products. Curr Opin Green Sustain Chem 2019;20:18–24. http://dx.doi.org/10.1016/j.cogsc.2019.08.001.

[6] Bhoi PR, et al. Recent advances on catalysts for improving hydrocarbon compounds in bio-oil of biomass catalytic pyrolysis. Renew Sustain Energy Rev 2020;121:109676. http://dx.doi.org/10.1016/j.rser.2019.109676.

[7] Andersen LF, et al. Biogas production from straw—The challenge feedstock pretreatment. Biomass Convers Biorefinery 2020;1–24. http://dx.doi.org/10.1007/s13399-020-00740-y.

[8] Yao Y, Shekhar DK. State of the art review on model predictive control (MPC) in Heating Ventilation and Air-conditioning (HVAC) field. Build Environ 2021;200:107952. http://dx.doi.org/10.1016/j.buildenv.2021.107952.

[9] Choudhari VG, Dhoble AS, Sathe TM. A review on effect of heat generation and various thermal management systems for lithium-ion battery used for electric vehicle. J Energy Storage 2020;32:101729. http://dx.doi.org/10.1016/j.est.2020.101729.

[10] Shivhare A, et al. Metal phosphate catalysts to upgrade lignocellulose biomass into value-added chemicals and biofuels. Green Chem 2021;23(11):3818–41. http://dx.doi.org/10.1039/D1GC00376C.

[11] Ehsanullah S, et al. How energy insecurity leads to energy poverty? Do environmental consideration and climate change concerns matter. Environ Sci Pollut Res 2021;28(39):55041–52. http://dx.doi.org/10.1007/s11356-021-14415-2.

[12] Kang J-N, et al. Energy systems for climate change mitigation: A systematic review. Appl Energy 2020;263:114602. http://dx.doi.org/10.1016/j.apenergy.2020.114602.

[13] Aslanturk O. The role of renewable energy in ensuring energy security of supply and reducing energy-related imports. Int J Energy Econ Policy 2020. http://dx.doi.org/10.32479/ijeep.8414.

[14] Doğan B, et al. Formulating energy security strategies for a sustainable environment: Evidence from the newly industrialized economies. Renew Sustain Energy Rev 2023;184:113551. http://dx.doi.org/10.1016/j.rser.2023.113551.

[15] Nyiwul L. Climate change adaptation and inequality in Africa: Case of water, energy and food insecurity. J Clean Prod 2021;278:123393. http://dx.doi.org/10.1016/j.jclepro.2020.123393.

[16] Hurlimann A, et al. Urban planning policy must do more to integrate climate change adaptation and mitigation actions. Land Use Policy 2021;101:105188. http://dx.doi.org/10.1016/j.landusepol.2020.105188.

[17] Ren X, et al. Challenges and opportunities in microwave-assisted catalytic pyrolysis of biomass: A review. Appl Energy 2022;315:118970. http://dx.doi.org/10.1016/j.apenergy.2022.118970.

[18] Liu C, et al. Microwave-assisted catalytic pyrolysis of apple wood to produce biochar: Co-pyrolysis behavior, pyrolysis kinetics analysis and evaluation of microbial carriers. Bioresour Technol 2021;320:124345. http://dx.doi.org/10.1016/j.biortech.2020.124345.

[19] Zhang T, et al. Microwave-assisted catalytic pyrolysis of waste printed circuit boards, and migration and distribution of bromine. J Hard Mater 2021;402:123749. http://dx.doi.org/10.1016/j.jhazmat.2020.123749.

[20] Zhang S, et al. Synthesis of CaO from waste shells for microwave-assisted catalytic pyrolysis of waste cooking oil to produce aromatic-rich bio-oil. Sci Total Environ 2022;827:154186. http://dx.doi.org/10.1016/j.scitotenv.2022.154186.

[21] Fodah AEM, Ghosal MK, Behera D. Bio-oil and biochar from microwave-assisted catalytic pyrolysis of corn stover using sodium carbonate catalyst. J Energy Inst 2021;94:242–51. http://dx.doi.org/10.1016/j.cej.2021.129412.

[22] Chen C, et al. Microwave-assisted catalytic pyrolysis of dunaliella salina using different compound additives. Renew Energy 2020;149:806–15. http://dx.doi.org/10.1016/j.renene.2019.12.089.

[23] Zhang Y, et al. Fast microwave-assisted pyrolysis of wastes for biofuels production–A review. Bioresour Technol 2020;297:122480. http://dx.doi.org/10.1016/j.biortech.2019.122480.

[24] Liu S, et al. Microwave-assisted metal-catalyzed pyrolysis of low-rank coal: Promising option towards obtaining high-quality products. J Energy Inst 2020;93(4):1602–14. http://dx.doi.org/10.1016/j.joei.2020.01.022.

[25] Anis S, et al. Microwave-assisted pyrolysis and distillation of cooking oils for liquid bio-fuel production. J Anal Appl Pyrolysis 2021;154:105014. http://dx.doi.org/10.1016/j.jaap.2020.105014.

[26] Narde SR, Remya N. Biochar production from agricultural biomass through microwave-assisted pyrolysis: predictive modelling and experimental validation of biochar yield. Environ Dev Sustain 2022;1–14. http://dx.doi.org/10.1007/s10668-021-01898-9.

[27] Siddique IJ, et al. Technical challenges in scaling up the microwave technology for biomass processing. Renew Sustain Energy Rev 2022;153:111767. http://dx.doi.org/10.1016/j.rser.2021.111767.

[28] Mishra RK, et al. Catalytic pyrolysis of biomass over zeolites for bio-oil and chemical production: A review on their structure, porosity, acidity co-relation. Bioresour Technol 2022;128189. http://dx.doi.org/10.1016/j.biortech.2022.128189.

[29] Valle B, et al. Role of zeolite properties in bio-oil deoxygenation and hydrocarbons production by catalytic cracking. Fuel Process Technol 2022;227:107130. http://dx.doi.org/10.1016/j.fuproc.2021.107130.

[30] Yadykova AY, Ilyin SO. Rheological and adhesive properties of nanocomposite bitumen binders based on hydrophilic or hydrophobic silica and modified with bio-oil. Constr Build Mater 2022;342:127946. http://dx.doi.org/10.1016/j.conbuildmat.2022.127946.

[31] Ly HV, et al. Catalytic pyrolysis of spent coffee waste for upgrading sustainable bio-oil in a bubbling fluidized-bed reactor: experimental and techno-economic analysis. Chem Eng J 2022;427:130956. http://dx.doi.org/10.1016/j.cej.2021.130956.

[32] Wu Q, et al. Microwave-assisted pyrolysis of waste cooking oil for hydrocarbon bio-oil over metal oxides and HZSM-5 catalysts. Energy Convers Manage 2020;220:113124. http://dx.doi.org/10.1016/j.enconman.2020.113124.

[33] Li J, et al. Machine learning aided bio-oil production with high energy recovery and low nitrogen content from hydrothermal liquefaction of biomass with experiment verification. Chem Eng J 2021;425:130649. http://dx.doi.org/10.1016/j.cej.2021.130649.
[34] Leng L, et al. Machine learning prediction of nitrogen heterocycles in bio-oil produced from hydrothermal liquefaction of biomass. Bioresour Technol 2022;362:127791. http://dx.doi.org/10.1016/j.biortech.2022.127791.
[35] Akbari M, Banitalebi-Dehkordi A, Zhang Y. E-lang: Energy-based joint inferencing of super and swift language models. 2022, http://dx.doi.org/10.48550/arXiv.2203.00748, arXiv preprint arXiv:2203.00748.
[36] Deng Y, et al. Residual energy-based models for text generation. 2020, http://dx.doi.org/10.48550/arXiv.2004.11714, arXiv preprint arXiv:2004.11714.
[37] Mireshghallah F, Goyal K, Berg-Kirkpatrick T. Mix and match: Learning-free controllable text generation using energy language models. 2022, http://dx.doi.org/10.48550/arXiv.2203.13299, arXiv preprint arXiv:2203.13299.
[38] Hoffmann J, et al. Training compute-optimal large language models. 2022, http://dx.doi.org/10.48550/arXiv.2203.15556, arXiv preprint arXiv:2203.15556.
[39] Kasneci E, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ 2023;103:102274. http://dx.doi.org/10.1016/j.lindif.2023.102274.
[40] Nugroho NA, Setiawan EB. Implementation Word2Vec for feature expansion in Twitter sentiment analysis. J RESTI (Rekayasa Sistem dan Tek Inf) 2021;5(5):837–42. http://dx.doi.org/10.29207/resti.v5i5.3325.
[41] Hong S, et al. Screening ideas in the early stages of technology development: A word2vec and convolutional neural network approach. Technovation 2022;112:102407. http://dx.doi.org/10.1016/j.technovation.2021.102407.
[42] Chandak A, Lee W, Stamp M. A comparison of word2vec, hmm2vec, and pca2vec for malware classification. In: Malware analysis using artificial intelligence and deep learning. 2021, p. 287–320. http://dx.doi.org/10.1007/978-3-030-62582-5_11.
[43] Wang H, et al. Neural-SEIR: A flexible data-driven framework for precise prediction of epidemic disease. Math Biosci Eng 2023;20(9):16807–23. http://dx.doi.org/10.3934/mbe.2023749.
[44] Spreafico C, Russo D, Degl'Innocenti R. Laser pyrolysis in papers and patents. J Intell Manuf 2022;1–33. http://dx.doi.org/10.1007/s10845-021-01809-9.
[45] Feng L, et al. Discovering technology opportunity by keyword-based patent analysis: a hybrid approach of morphology analysis and USIT. Sustainability 2019;12(1):136. http://dx.doi.org/10.3390/su12010136.
[46] Deng G, et al. Jailbreaker: Automated jailbreak across multiple large language model chatbots. 2023, http://dx.doi.org/10.48550/arXiv.2307.08715, arXiv preprint arXiv:2307.08715.
[47] Liu J, et al. Biomass pyrolysis technology by catalytic fast pyrolysis, catalytic co-pyrolysis and microwave-assisted pyrolysis: A review. Catalysts 2020;10(7):742. http://dx.doi.org/10.3390/catal10070742.
[48] Mohabeer C, et al. Microwave-assisted pyrolysis of biomass with and without use of catalyst in a fluidised bed reactor: a review. Energies 2022;15(9):3258. http://dx.doi.org/10.3390/en15093258.
[49] Yu Y, et al. Influence of catalyst types on the microwave-induced pyrolysis of sewage sludge. J Anal Appl Pyrolysis 2014;106:86–91. http://dx.doi.org/10.1016/j.jaap.2014.01.003.
[50] Mushtaq F, Mat R, Ani FN. A review on microwave-assisted pyrolysis of coal and biomass for fuel production. Renew Sustain Energy Rev 2014;39:555–74. http://dx.doi.org/10.1016/j.rser.2014.07.073.
[51] Zhang B, et al. Microwave-assisted catalytic fast pyrolysis of spent edible mushroom substrate for bio-oil production using surface modified zeolite catalyst. J Anal Appl Pyrolysis 2017;123:92–8. http://dx.doi.org/10.1016/j.jaap.2016.12.022.
[52] Latrasse L, et al. Self-matching plasma sources using 2.45 GHz solid-state generators: microwave design and operating performance. J Microw Power Electromagn Energy 2017;51(4):237–58. http://dx.doi.org/10.1080/08327823.2017.1388338.
[53] Latrasse L, et al. 2.45-GHz microwave plasma sources using solid-state microwave generators. Collisional-type plasma source. J Microw Power Electromagn Energy 2017;51(1):43–58. http://dx.doi.org/10.1080/08327823.2017.1293589.
[54] Mayerhöfer TG, et al. Consolidated silica glass from nanoparticles. J Solid State Chem 2008;181(9):2442–7. http://dx.doi.org/10.1016/j.jssc.2008.06.011.
[55] Weber JR, et al. Glass fibres of pure and erbium-or neodymium-doped yttria–alumina compositions. Nature 1998;393(6687):769–71. http://dx.doi.org/10.1038/31662.
[56] McDonald E, Zmeureanu R. Virtual flow meter for chilled and condenser water for chillers: Estimates versus measurements. Sci Technol Built Environ 2016;22(2):178–88. http://dx.doi.org/10.1080/23744731.2015.1085279.
[57] Xu Y, et al. Research the wet gas flow measurement based on dual-throttle device. Flow Meas Instrum 2013;34:68–75. http://dx.doi.org/10.1016/j.flowmeasinst.2013.07.014.
[58] Mo S, et al. Passive control of gas–liquid flow in a separator unit using an apertured baffle in a parallel-flow condenser. Exp Therm Fluid Sci 2014;53:127–35. http://dx.doi.org/10.1016/j.expthermflusci.2013.11.017.
[59] Zou J, Han Y, So S-S. Overview of artificial neural networks. In: Artificial neural networks: Methods and applications. 2009, p. 14–22. http://dx.doi.org/10.1007/978-1-60327-101-1_2.
[60] Grohe M. word2vec, node2vec, graph2vec, x2vec: Towards a theory of vector embeddings of structured data. In: Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI symposium on principles of database systems. 2020, http://dx.doi.org/10.1145/3375395.3387641.
[61] Liu Y, et al. Summary of chatgpt-related research and perspective towards the future of large language models. Meta-Radiol 2023;100017. http://dx.doi.org/10.1016/j.metrad.2023.100017.
[62] Li B, et al. Application of artificial neural networks to photovoltaic fault detection and diagnosis: A review. Renew Sustain Energy Rev 2021;138:110512. http://dx.doi.org/10.1016/j.rser.2020.110512.
[63] Floridi L, Chiriatti M. GPT-3: Its nature, scope, limits, and consequences. Minds Mach 2020;30:681–94. http://dx.doi.org/10.1007/s11023-020-09548-1.
[64] Chen M, et al. Evaluating large language models trained on code. 2021, http://dx.doi.org/10.48550/arXiv.2107.03374, arXiv preprint arXiv:2107.03374.
[65] Di Gennaro G, Buonanno A, Palmieri FA. Considerations about learning Word2Vec. J Supercomput 2021;1–16. http://dx.doi.org/10.1007/s11227-021-03743-2.
[66] Fan F-L, et al. On interpretability of artificial neural networks: A survey. IEEE Trans Radiat Plasma Med Sci 2021;5(6):741–60. http://dx.doi.org/10.1109/TRPMS.2021.3066428.
[67] Zhang F, O'Donnell LJ. Support vector regression. In: Machine learning. Academic Press; 2020, p. 123–40. http://dx.doi.org/10.1016/B978-0-12-815739-8.00007-9.
[68] Sekulić A, et al. Random forest spatial interpolation. Remote Sens 2020;12(10):1687. http://dx.doi.org/10.3390/rs12101687.
[69] Yang Y, et al. Biomass microwave pyrolysis characterization by machine learning for sustainable rural biorefineries. Renew Energy 2022;201:70–86. http://dx.doi.org/10.1016/j.renene.2022.11.028.
[70] Nie P, et al. Prediction of home energy consumption based on gradient boosting regression tree. Energy Rep 2021;7:1246–55. http://dx.doi.org/10.1016/j.egyr.2021.02.006.
[71] Ascher S, Watson I, You S. Machine learning methods for modelling the gasification and pyrolysis of biomass and waste. Renew Sustain Energy Rev 2022;155:111902. http://dx.doi.org/10.1016/j.rser.2021.111902.
[72] Mari Selvam S, Balasubramanian P. Influence of biomass composition and microwave pyrolysis conditions on biochar yield and its properties: a machine learning approach. BioEnergy Res 2023;16(1):138–50. http://dx.doi.org/10.1007/s12155-022-10447-9.
[73] Dong Z, et al. Machine learning prediction of pyrolytic products of lignocellulosic biomass based on physicochemical characteristics and pyrolysis conditions. Bioresour Technol 2023;367:128182. http://dx.doi.org/10.1016/j.biortech.2022.128182.