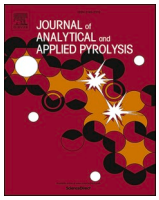
Journal of Analytical and Applied Pyrolysis 169 (2023) 105857

Contents lists available at ScienceDirect

Journal of Analytical and Applied Pyrolysis

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



Applied machine learning for prediction of waste plastic pyrolysis towards valuable fuel and chemicals production

Yi Cheng [a,b], Ecrin Ekici [c], Güray Yildiz [c], Yang Yang [a,b], Brad Coward [a], Jiawei Wang [a,b,*]

[a] Department of Chemical Engineering and Applied Chemistry, Aston University, Birmingham B4 7ET, UK
[b] Energy and Bioproducts Research Institute (EBRI), Aston University, Birmingham B4 7ET, UK
[c] Department of Energy Systems Engineering, Faculty of Engineering, Izmir Institute of Technology, Urla 35430, Izmir, Türkiye

ARTICLE INFO

Keywords:
Waste plastics
Pyrolysis
Machine learning
Decision tree
Ultimate analysis

ABSTRACT

Pyrolysis is a suitable conversion technology to address the severe ecological and environmental hurdles caused by waste plastics' ineffective pre- and/or post-user management and massive landfilling. By using machine learning (ML) algorithms, the present study developed models for predicting the products of continuous and non-catalytically processes for the pyrolysis of waste plastics. Along with different input datasets, four algorithms, including decision tree (DT), artificial neuron network (ANN), support vector machine (SVM), and Gaussian process (GP), were compared to select input variables for the most accurate models. Among these algorithms, the DT model exhibited generalisable and satisfactory accuracy ($R^2 > 0.99$) with training data. The dataset with the elemental composition of waste plastics achieved better accuracy than that with the plastic-type for predicting liquid yields. These observations allow the predictions by the data from ultimate analysis when inaccessible to the plastic-type data in unknown plastic wastes. Besides, the combination of ultimate analysis input and the DT model also achieved excellent accuracy in liquid and gas composition predictions.

1. Introduction

Plastic plays a vital role in modern society and is a non-negligible part of the construction, healthcare, electronics, automotive, and packaging industries [1]. Since their first appearance, the production of plastics has skyrocketed with the growing global population and social demands [2]. Most plastics are petroleum-based polymers, such as polyethylene (PE), polypropylene (PP), polystyrene (PS), polyvinyl chloride (PVC), and polyethylene terephthalate (PET) [3]. Therefore, plastic production highly depends on fossil resources, making them unsustainable without proper recycling pathways.

However, existing waste management systems worldwide lack sufficient capacity at the global level to dispose of or recycle all waste plastics safely. It resulted in an inevitable increase in waste plastic disposal in the environment [4]. Previous studies estimated that 8 million metric tons of macro-plastic and 1.5 million metric tons of primary micro-plastic enter the ocean annually [5]. Waste plastics may take up to billions of years to degrade in the ecosystem [6]. The alternative to landfilling is mechanical recycling, chemical recycling, and energy recovery. Mechanical recycling of plastic waste is the reprocessing of plastic waste into new and serviceable materials. However, the drawback is the costly and inefficient pre-separation by labour. The pre-separation is inevitable, as waste plastics have different resins, transparency, and colour. Mixed plastics are undesirable for manufacturers because mixtures have fewer transform abilities and lower flexibilities [7]. In addition, pre-separation consumes large amounts of water for cleaning, and the resulting water contamination reduces the sustainability of the recovery.

Chemical recycling of waste plastics is more practicable, as most industries manufacture plastics from fossil fuels. It is an environmentally friendly way to meet the increased energy demand through waste recycling [8]. Pyrolysis is a method of chemical recycling to degrade long-chain polymer molecules into smaller, less complex molecules in an inert atmosphere. In the past two decades, extensive research and development work promoted the technological development of the pyrolysis of waste plastic [9,10]. The process requires intense heat with a shorter duration in an oxygen-free atmosphere and generates initial volatiles and solid residue [11]. Subsequently, initial volatiles condensed and further formatted the liquid and non-condensable gas. The liquid products have multiple applications in furnaces, boilers, turbines, and diesel engines, without upgrading or further treatment. The by-product non-condensable gas has a substantial calorific value

* Corresponding author at: Department of Chemical Engineering and Applied Chemistry, Aston University, Birmingham B4 7ET, UK.
E-mail address: j.wang23@aston.ac.uk (J. Wang).

https://doi.org/10.1016/j.jaap.2023.105857
Received 26 November 2022; Received in revised form 21 December 2022; Accepted 2 January 2023
Available online 5 January 2023

and compensates for balancing the overall energy of the process [12].

The pyrolysis process optimisation could achieve different product yields and distributions by manipulating the operating parameters. Previous studies have recognised that liquid product yields and quality depend on the pyrolysis parameters, which involve the feedstock properties, pyrolysis temperature, reactor type, residence times of plastics and primary pyrolysis vapours, pressure, and carrier gas with its flow rate [13,14]. Different plastics contain various amounts of volatile matter and ash content. In most cases, a higher volatile matter favours liquid conversion [15]. For instance, the pyrolysis of PP, HDPE (high--density polyethylene), and LDPE (low-density polyethylene) yielded 80.1 wt%, 84.7 wt%, and 93.1 wt% liquid products at their optimised reaction temperatures, respectively [16,17]. The pyrolysis of PS has an even higher liquid oil yield of 97.6 wt% at the optimum temperature of 425 °C [18]. Unlike polyolefins (PP and PE) and PS, other common plastics, such as PET and PVC, lead to a lower liquid product yield. The PET pyrolysis produced only 23.1 wt% oil and 76.9 wt% gas when conducting the reactor at about 500 °C [19]. In the pyrolysis of PVC in a batch reactor at 520 °C with a heating rate of 10 °C /min, the liquid product yield was about 12.8 wt%, of which hydrogen chloride content was up to 58 wt% [20]. These studies showed that each type of plastic has specialised pyrolysis characteristics and behaviours. However, waste plastics are often a mixture of various plastics, and it is difficult, if not impossible, to get accurate data on compositions. The pyrolysis of mixed plastics often produced a lower liquid yield of less than 50 wt% in experimental results [8]. Therefore, optimising the pyrolysis parameters based on complicated feedstock still requires attention to develop the process further, maximise liquid production and improve the quality.

Process modelling is a promising tool for addressing complicated systems and is necessary for the scale-up and optimization of industrial processes. Because of the heterogeneity of the physicochemical structure of the plastic mixture, the development of mathematical models to simulate plastic mixture pyrolysis requires a topology algorithm in most cases [21]. Machine learning (ML) successfully predicted the pyrolysis of biomass, coal, and organic solid waste [22–24]. The ML can improve automatically through data uptake and experience, attempting to map inputs to the corresponding responses to comprehend the mathematical linkages from various complex processes with algorithms [22]. Several algorithms have modelled pyrolysis or gasification processes in recent reports, including artificial neural networks (ANN), tree-based algorithms, like decision tree (DT) and random forest, and support vector machine (SVM) [25,26]. ANN method is the ML technique receiving the most attention. For instance, when modelling biomass gasification by featuring tar, char, and permanent gas interactions, ANN outperformed a real-gas equilibrium model for predicting the gasification products [27]. Beyond ANN, other relevant ML options, such as the tree-based approach and SVM, have some successful implementations. Cheng et al. proposed integrating an RF-based predictive model with life cycle assessment and economic analysis [28]. They aimed to achieve a comprehensive evaluation holistically with different pyrolysis feedstocks. SVM also has a wide application in pyrolysis prediction tasks. For example, in predicting pyrolysis biochar yield, SVM showed better performance than ANN at R^2 and RMSE [29].

To our best knowledge, the research on waste plastic pyrolysis is still short for easy, cheap, and reliable methods for complex feedstock characterisation. Corresponding data absence limited the possibilities of yield maximisation by optimising the operating parameters, and product prediction is challenging when facing the complexity of feedstock that varies in each batch. This study aims to model the relationships between the input variables, such as the feedstock and operating parameters, of a non-catalytically operating pyrolysis process and the responses, which include the three-phase product yields and chemicals in them. By comparing the algorithms of the DT, ANN, SVM, and Gaussian process (GP), this work clarified practical and operable operating parameters for identifying the most accurate models. The models developed in this work showed the potential to predict and optimise the non-catalytic pyrolysis process of waste plastic without expensive experiments.

2. Material and methods

2.1. Data collection

This study reviewed 93 relevant works (Table S1 and S2) to get experimental data and develop the prediction models. The research articles were collected using Web of Science, Scopus, and Wiley databases. The collection comprised articles written in English and published between January 1, 1984, and December 31, 2021.

The keywords used for the database search were “pyrolysis”, “waste plastic”, “polyethylene”, “polypropylene”, “polystyrene”, “polyvinylchloride”, and “polyethylene terephthalate”. Articles matching these keywords are collected and classified into two sets: *review articles* and *research articles*. As a supportive step, the individual lists of references of the review articles were investigated further to expand the collected set of research articles. In line with the focus of this work, only the articles reporting the results of “continuously” and “non-catalytically” operating pyrolysis systems were added to the collection. Research articles reporting the results both for “non-catalytic” and “catalytic” tests were also considered. However, only the results involving the “non-catalytic” operations were considered.

The input variables in the study included the feedstock properties, such as compositions of different plastic types (wt%), ultimate analysis (wt%, ash free) and particle size (mm). The plastic-type included in the study were polyethylene (PE), polypropylene (PP), polystyrene (PS), polyvinyl chloride (PVC), and polyethylene terephthalate (PET). The ash-free elemental compositions of carbon (C), hydrogen (H), oxygen (O), nitrogen (N), and chlorine (Cl) were included as the inputs, while sulphur was not selected as it was often negligible compared to the elements mentioned above. The feed intake capacity (kg/h), pyrolysis temperature (°C) and vapour residence time (s) were considered as the reactor operating parameters. Heating rate and carrier gas flow rate were not selected as input variables because only a small number of the selected references reported them. The mean value was used in the study for the input variables if a range was reported in the references.

Three datasets were compiled based on different selections of the input variables on feedstock properties. They were used to investigate the effects of the selections of the input variables on the predictability and accuracy of the developed models. Dataset 1 contained the compositions of different plastic types as the feedstock properties, and Dataset 2 used the results from ultimate analysis. In Dataset 3, both the compositions of plastic types and the ultimate analysis data were selected as input variables. The inputs for the reactor operating conditions were the same in all three datasets. Table 1 summarises the range of the input variables and the total numbers of the data points in each dataset.

The histograms in Fig. 1 showed the distributions of feedstock properties and reactor operating parameters used in the study. First, the plastic types had narrow distributions because the feedstocks used in laboratory-scale research were often single type plastics. In some studies that used real-world wastes as the feedstock, the compositions of different types of plastics were often unknown and therefore had not been included in the histogram. Secondly, the histograms of the ultimate analysis showed the common elements of C and H multi-distribution and O, N, and Cl single distribution. The reactor parameters of feed intake, pyrolysis temperature, and vapour residence time (VRT) ranged in multi-distribution. Most of the feed intake was less than 1 kg/h, the pyrolysis temperature was in the range of 400–800 °C, and the VRT of most reported research was around 1 s

2.2. Responses for the model development

Pyrolysis of plastics produces three main products. The liquid product are condensed volatiles (including wax), the gas product is the

Table 1
Descriptive statistics of input variable and data point amounts.

Inputs			Dataset 1	Dataset 2	Dataset 3
Feedstock	Plastic type	PE (wt %)	0–100	–	0–100
		PP (wt %)	0–100	–	0–100
		PS (wt%)	0–100	–	0–100
		PVC (wt %)	0–100	–	0–100
	Ultimate analysis	C (wt a.f %)	–	38.4–92.3	38.4–92.3
		H (wt a.f %)	–	4.21–14.5	4.21–14.5
		N (wt a.f %)	–	0–6.3	0–6.44
		O (wt a.f %)	–	0–32.95	0–32.95
		Cl (wt a.f %)	–	0–56.8	0–56.8
	Particle size (mm)		0.05–5	0.05–5	0.05–5
Reactor	Feed intake (kg/h)		0.003–50	0.003–10.8	0.003–10.8
	Pyrolysis temperature (°C)		400–860	400–860	400–860
	Vapour residence time (s)		0–15	0–15	0–15
Total data points			275	288	274

a.f: ash-free.

non-condensable part, and the solid product are carbonaceous residue after the pyrolysis process. The liquid and gas products can be grouped into sub-products by their carbon number: C1–C4 gases, gasoline ranging from C5 to C12, diesel ranging from C13 to C20, and wax over C21. In addition, diesel contained styrene and BTX based on the petroleum classification method. Therefore, the responses selected in the study included overall yields (wt%) of liquid, gas, and solid products and the yields of the compounds in the liquid product, such as wax (> C20), aromatics, benzene-toluene-xylene (BTX), styrene, gasoline (C5–C12) and diesel (C13–20), and the yields (wt%) of C1-C4 and traced gases in the gas product. In the data collection, notably, BTX refers to the reaction of which targets were producing the BTX. Aromatics refers to all aromatic products which contain BTX in most cases and without specific focus.

For all responses, the following equations were used to normalise the data.

$$Y_{liquid} + Y_{gas} + Y_{solid} = 1 \quad (1)$$

$$Y_{wax} + Y_{diesel} + Y_{gasoline} = Y_{liquid} \quad (2)$$

$$Y_{C1} + Y_{C2} + Y_{C3} + Y_{C4} + Y_{tracegases} = Y_{gas} \quad (3)$$

where Y_x are the yields of compound x.

The histograms in Fig. 2 showed the distribution of the responses. The yields of liquid and gas products have an even distribution from 0 % to 100 %, while most of the references did not report the yield of the solid product as no solid residue was produced when a pure polymer was used in a pyrolysis process. In addition, the middle bar of Fig. 2(a) presented upper line data exceeding 100 % a little. It means some data involved non-vacuum pyrolysis. However, as the dominant distribution ranged under 70 %, it enabled normalising the data with Eq. (1) to conduct the dataset.

2.3. Algorithm methods comparison

Four algorithms were employed to develop the prediction models,

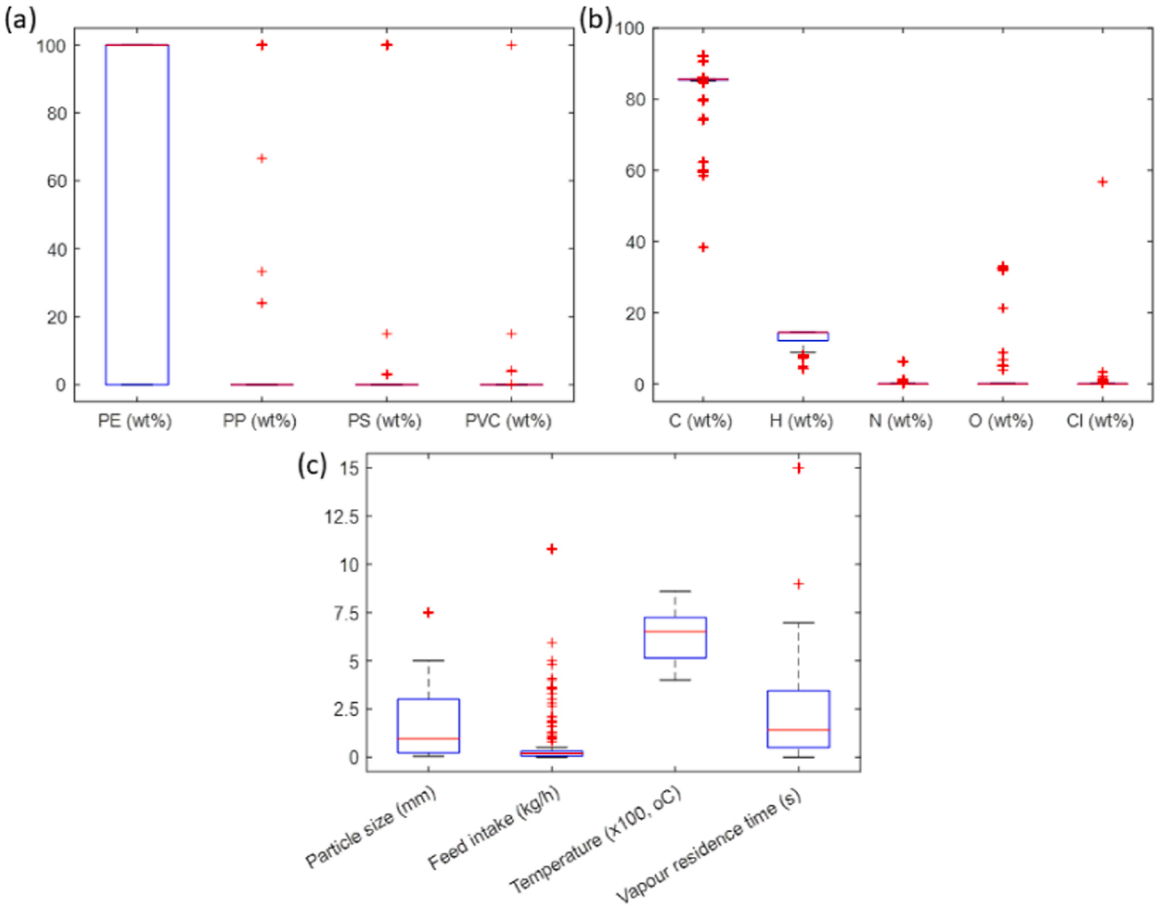


Fig. 1. Statistics of the input variables of plastic types (a), ultimate analysis (b), and operating parameters (c).

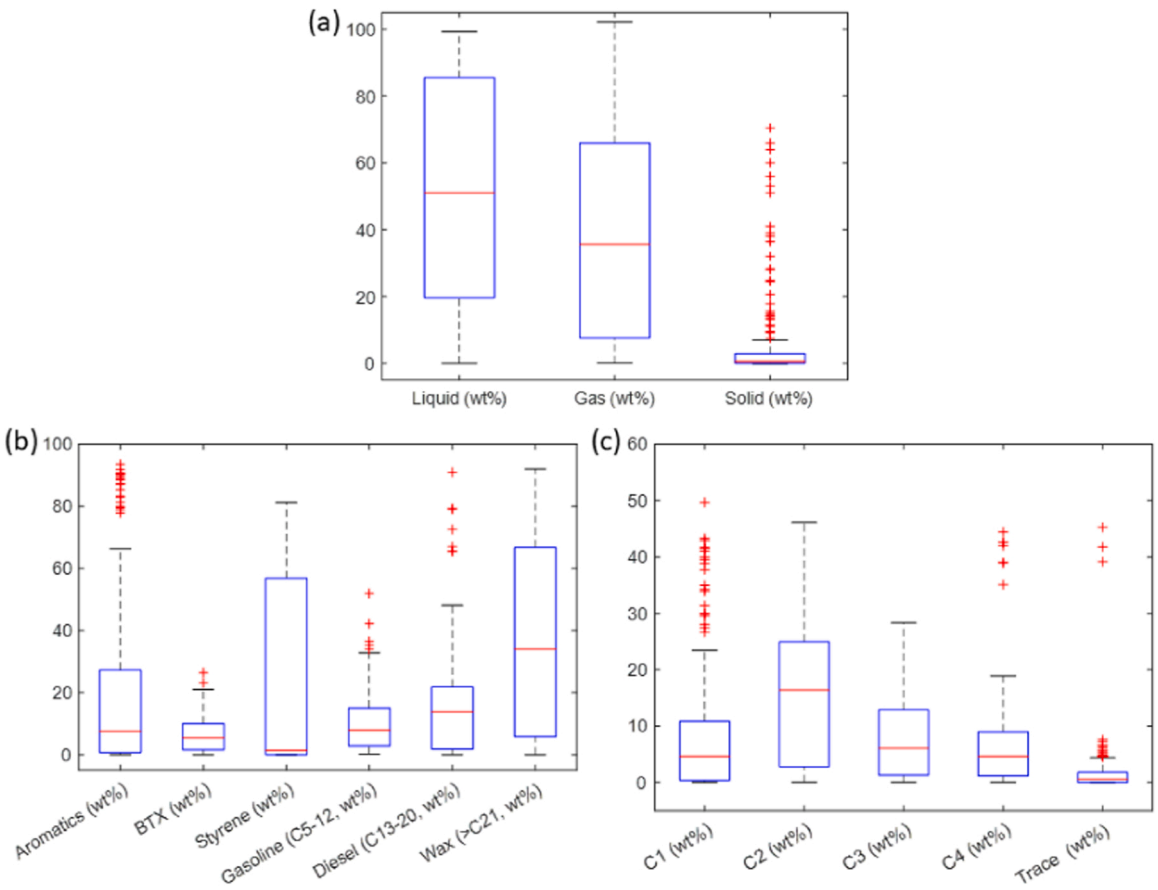


Fig. 2. Statistics of the responses of total mass balance (a), liquid compositions (b), and gas compositions (c).

including the decision tree, artificial neural networks, support vector machine, and Gaussian process. The data sets were divided into a training set (containing 70 % of total data points) and a validation-test set (containing 30 % of total data points). The division was randomly selected from the datasets to test the robustness and predictability of the models.

2.3.1. Decision tree (DT)

A decision tree (DT) algorithm is a supervised ML method, which generates a set of decision rules by repeatedly splitting the input dataset into binary sections. The search method for all input features examines the division until a homogeneous response distribution, and the hyper-parameter optimisation can be applied to develop the DT model further. In addition, the DT algorithm can use pruning modular to prevent over-fitting, which is the common drawback of the regression algorithms [30]. In this work, we used the function of *fitrensemble* in MATLAB to develop DT models.

2.3.2. Artificial neural networks (ANN)

An artificial neural network is a system based on the operation of biological neural networks, which analyses datasets and trains itself to recognise patterns between the input variables and responses. The basic unit of the network is the artificial neuron, which involves input, output, and several hidden layers (one for a single layer perception). Each neuron transmits received signals by interconnecting. The hidden layer connects to the input and target parameters by adjustable weighted linkages, and the transfer function in the hidden layer introduces nonlinearity to the network. Each layer has a weight matrix, a bias vector, and an output vector. The network learns by feeding back its predictions, comparing them to the corresponding inputs, and adjusting weights accordingly [31]. In this work, we used the function of *fitrnet* in MATLAB to develop ANN models.

2.3.3. Support vector machine (SVM)

The SVM is also a supervised learning model which constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space. The SVM uses the kernel function in classification and regression problems to develop the linear relationship between input and response. This algorithm uses the principle of the statistical structure risk minimisation to reduce the confidence range and get a small real risk. In the presence of noise, if the hyper-plane can still classify the samples well, then the classification can be the best when the distance between the sample points and the hyper-plane is the largest. SVM modelling is an algorithm that transforms the linear programming problem of solving the dual problem [29,32]. In this work, we used the function of *fitrsvm* in MATLAB to develop SWM models.

2.3.4. Gaussian process (GP)

The Gaussian process is a stochastic process by collecting random variables, of which every finite collection of those random variables has a multivariate normal distribution. The GP is one of the alternative approaches to neural networks since a large class of Bayesian regression models converged based on a neural network and is limited by an infinite Network. The reduction of the dimension of the variables is often desirable in developing a GP model for large amounts of covariate regression to ease the computational burden and improve the prediction accuracy. Variable selection performs better interpretability than the projection techniques, as the covariates are the most informative for prediction [33]. In this work, we used the function of *fitrgp* in MATLAB to develop GP models.

The coefficient of determination (R^2) and root-mean-square error (RMSE) were used to evaluate the prediction performance. The formulas

of R^2 and RMSE are shown as Eqs. (4) and (5):

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \bar{y}_n)^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{N}} \quad (5)$$

where N is the number of data points; y_n is the n^{th} data point, and $\hat{y}_n$ is the corresponding predicted value; $\bar{y}_n$ is the average value of the N data points.

3. Results and discussion

3.1. Correlation analysis of input variables and responses

Fig. 3 illustrates the correlation matrixes of each input variable and response. As shown in Fig. 3(a), the pyrolysis temperature strongly correlated with the responses. The liquid and gas yields have correlations of − 0.64 and 0.75 with the pyrolysis temperature, respectively. It indicated that plastic pyrolysis generated more gas and less liquid at higher temperatures. Likewise, the wax and styrene in the liquid product gave respective negative correlations of − 0.56 and − 0.61, and the gaseous components from C1 to C4 all gave positive correlations of over 0.5 with the pyrolysis temperature. In addition, the correlations between the plastic-type and the yields of solid, liquid, and gas products were not conclusive. Only in the liquid components, PE and PS have respective negative and positive correlations with the aromatic and styrene yields. Besides, carbon and hydrogen inputs also exhibited correlations with the yields of aromatic and styrene compounds. It may be because of high carbon-hydrogen ratio implied more unsaturated bonds, such as aromatic rings, in the polymers. Particle size has a weak positive correlation with the yield of solids and diesel. A bigger particle size could lead to slower mass and heat transfer, hence more solid residue at the end of the reaction.

Pearson correlation matrix of the inputs in Fig. 3(b) provided more information about the relationship between the input variables. In terms of the feedstock ultimate analysis results, hydrogen correlated 0.69 and − 0.50 with PE and PS, respectively. It might be because the hydrogen content reflected the unsaturated number of the polymer in some contexts. The chlorine content performed the highest correlation of 0.99 with PVC, which showed that PVC was the only source of chlorine in the selected samples. Oxygen showed obvious negative correlations with carbon and hydrogen. It might be because the oxygen content accounts for a high proportion of some oxygenated polymers, such as PET. Besides, vapour residue time (VRT) and intake showed a 0.71 correlation, which confirmed that feeding mode played a vital role in plastic pyrolysis. Larger operations typically have longer vapour residue time. Overall, the correlation matrix of inputs implied some connections existed between the input variables, which might overcome the shortages of input information.

3.2. Screening of the algorithms and datasets

Four algorithms, ANN, GP, SVM, and DT, were evaluated in this study for their applicability. The response used in the evaluation was the yield of liquid product, and the criteria were the coefficient of determination (R^2), and the root-mean-squared error (RMSE). Table 2 (entries 1–4) shows the R^2 and RMSE values for the ANN, GP, SVM, and DT algorithms and Dataset 1. Fig. 4(a–d) were parity plots displaying response measures with R^2 values for four ML models.

Fig. 4 and Table 2 showed that the DT algorithm predicted the response with the best accuracy among the four algorithms. The training sets and the DT compensated for the tendency of overfitting [34]. The training data showed an R^2 value of 0.984, and the testing data was 0.882. The second-best algorithm was ANN, which predicted the response with an R^2 value of 0.926 for the training data and 0.719 for the testing data. The RMSE values for the training and testing sets based on Datasets 1 were also available in Table 2. The DT showed a smaller difference in RMSEs between the training and testing sets, and the ANN gave higher RMSE than the DT for both training and test sets. The GP algorithm got an R^2 value of 0.592 for the training set and 0.534 for the testing set, and the SVM algorithm achieved an R^2 of 0.583 and 0.653 for the training and testing sets, respectively. At the same time, the RMSEs of the GP and SVM models were higher than those of the DT and ANN models on both training and testing sets. Hence, the GP and SVM algorithms were unsuitable for predicting the yield of the liquid product in this study.

In early reports, tree-based algorithms, such as ensemble methods, have been less frequently used than ANN in pyrolysis predictions [22]. However, in a recent comparison between the ANN and tree-based algorithms for pyrolysis composition prediction, both have achieved similar prediction accuracy [35]. Although the ensemble methods of DT suffer from a similar black-box nature as ANN, the interpretability of tree-based methods exceeds ANN, as the feature importance was easier to extract [30,32]. Besides, the choice of a machine learning model depends on the level of accuracy required and the amount of training data available [36]. Hence, the DT was the best choice if adequate training data was available.

Table 2 and Fig. 4 also illustrate the comparative evaluation of the choice of different datasets using the same algorithm. As mentioned in Section 2.2, Dataset 1 comprised plastic-type and particle size as the feedstock properties, and in Dataset 2, the elemental composition data replaced plastic-type. Dataset 3 contained both plastic-type and ultimate

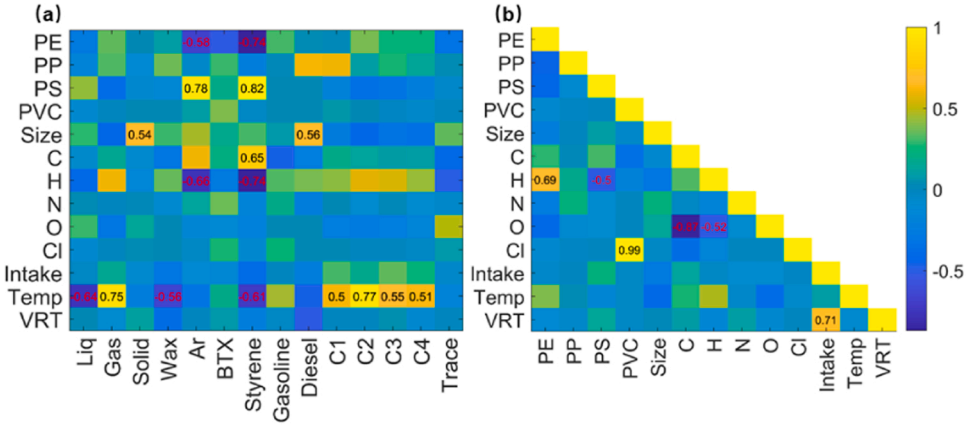


Fig. 3. Pearson correlation coefficients between the input variables and responses and between the inputs and inputs.

Table 2
Summary of each algorithm with the prediction of the liquid yield.

Run	Dataset	Data points	Method	Training		Testing	
				R^2	RMSE	R^2	RMSE
1	1	275	ANN (fitrnet)	0.926	9.6	0.719	18.1
2	1	275	GP (fitrgp)	0.592	22.4	0.534	23.4
3	1	275	SVM (fitrsvm)	0.583	22.2	0.653	21.0
4	1	275	DT (fitrensemble)	0.984	4.3	0.882	11.7
5	2	301	DT (fitrensemble)	0.996	2.2	0.869	12.5
6	3	275	DT (fitrensemble)	0.987	3.8	0.859	13.4

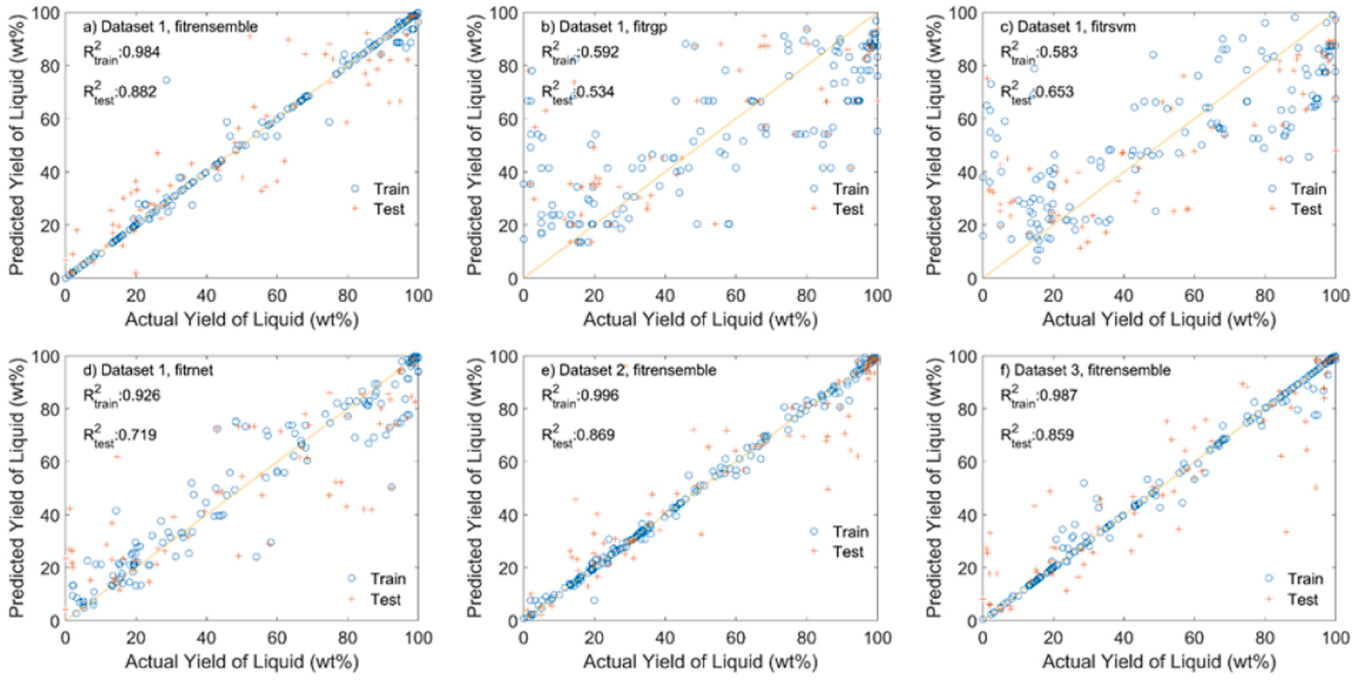


Fig. 4. Scatter plots of the pairs of input variables and responses with different algorithms and datasets: (a) Dataset 1 with fitrensemble, (b) Dataset 1 with fitrgp, (c) Dataset 1 with fitrsvm, (d) Dataset 1 with fitrnet, (e) Dataset 2 with fitrensemble, and (f) Dataset 3 with fitrensemble.

analysis data. The numbers of the data points for the three datasets were 275, 301, and 275, of which Datasets 1 and 3 did not share the same sets. In the training phase, Dataset 2 had the highest R^2 value of 0.996, followed by Dataset 3 of 0.987 and Dataset 1 of 0.984. Likewise, Dataset 2 had the lowest RMSE of 2.2, lower than Dataset 3 (3.8) and Dataset 1 (4.3). These results showed that in the cases of ash-free plastics pyrolysis, the elemental compositions might be better as input variables than the plastic type. This result overturned the previous perception that tracing plastic-type was the foremost in product prediction. Furthermore, it is challenging to accurately identify plastic types in mixtures with inefficient sorting systems in recycling waste plastic [37]. As shown in the correlation matrix in Fig. 3, the ultimate analysis results had a certain degree of correlation with plastic-type in plastic mixtures. Most polymers have their specific monomer structures. Apart from the end area, those structures of monomer linking have close dissociation energies, implying that heating up to the dissociation level might be the key to mixed plastic non-catalytical pyrolysis modelling [38]. This finding is significant as it is easier to characterize plastic mixtures with elemental compositions and then use these results to predict the pyrolytic products of plastics.

3.3. Total mass balance predictions

Table 3 shows the R^2 and RMSE of the responses with Dataset 2 for estimated training and testing data using fitrensemble. The total balance of final products, i.e. liquid, gas and solid, was 100 %. The R^2 values for training data of the yields of the three products were all over 0.995, and the ones for testing data were slightly lower. Meanwhile, Fig. 5 gives the graphical presentation of the results and the predictor importance estimate for every target. The primary feature for liquid and gas yields predictions was the pyrolysis temperature, and the hydrogen content and particle size were after temperature. It showed that the feature importance gained from DT was the same as the prediction by using correlation coefficients. Besides, feeding intake was most important, followed by carbon content and particle size for solid or char prediction. However, compared with similar works of biomass pyrolysis, it received opposite results when using the DT algorithm. The biomass composition has more effect than the pyrolysis conditions on the liquid yield response [39]. Researchers found that pyrolysis temperature was the most important factor in biochar production [28,37]. Biomass, especially for lignocellulose, has the structure of net-like linkage and high oxygen content, which is perhaps easier than the line structure of plastics to format the char. Polymer type, therefore, might play an essential role in pyrolysis. Reconsideration was necessary for the relationships between feedstock compositions and operating conditions, especially in the co-pyrolysis process of biomass and plastics [40–43].

3.4. Compositions of liquid and gas predictions

Pyrolysis liquids are the condensed parts of pyrolytic products, including wax, diesel, and gasoline. Here we viewed the aromatics as the composition of the diesel and outside the liquid mass balance. As

Table 3
Summary of target predictions with decision tree (DT) algorithm.

Category	Target	Number of non-zero data points	Training		Testing	
			R^2	RMSE	R^2	RMSE
Total mass balance	Liquid yield	276	0.997	1.8	0.940	7.8
	Gas yield	287	0.997	1.7	0.914	9.1
	Solid yield	139	0.995	0.9	0.845	5.3
Liquid compositions	Wax	106	0.997	1.7	0.786	14.5
	Aromatics	166	0.997	1.8	0.898	8.8
	Gasoline	112	0.991	2.6	0.899	7.4
	Diesel	91	0.906	2.9	0.803	4.4
Gas compositions	C1	229	0.981	1.6	0.616	5.7
	C2	221	0.991	1.3	0.842	5.6
	C3	225	0.983	1.0	0.633	4.6
	C4	215	0.991	0.7	0.742	4.2
	Trace gas	124	0.992	0.5	0.892	1.7

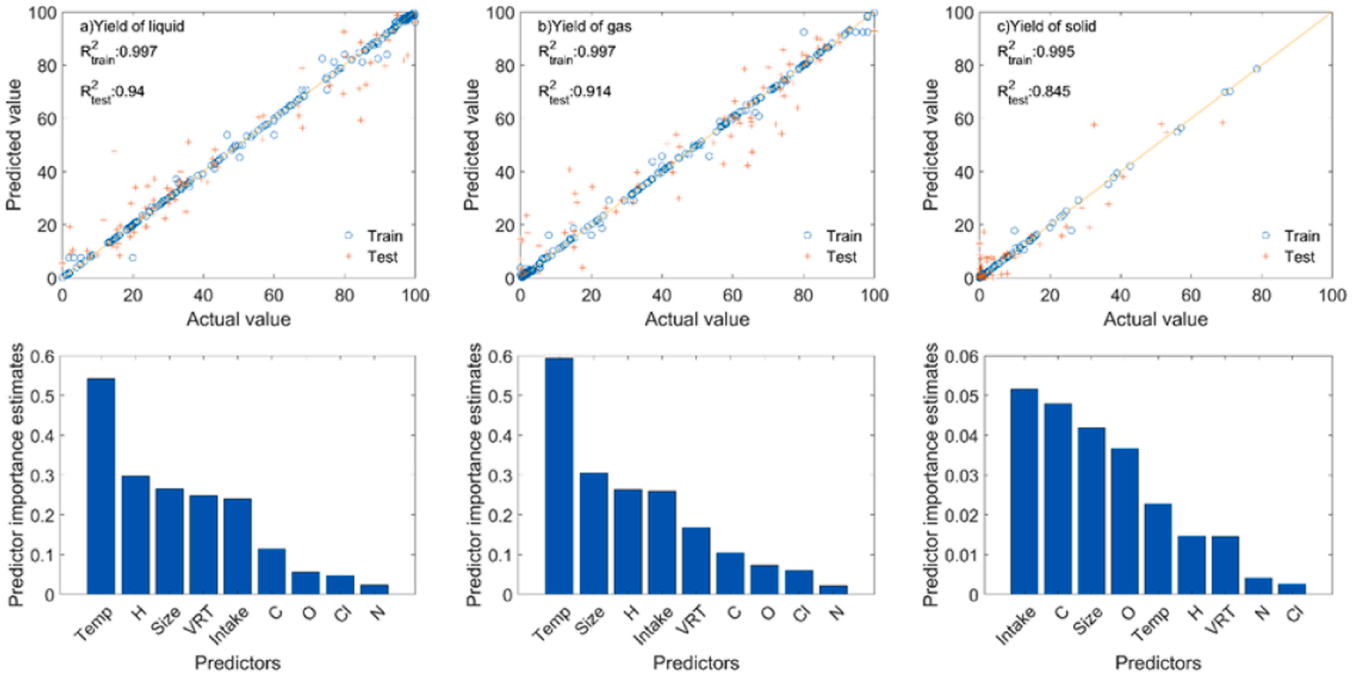


Fig. 5. Plots of the actual and predicted values obtained from the fitrensemble with Dataset 2 for total mass balance: (a) yield of liquid; (b) yield of gas, and (c) yield of solid.

depicted in Table 3, wax and aromatics showed the highest R^2 values of 0.997, followed by gasoline of 0.991 and diesel of 0.906 among the training data. Another notable result was that the R^2 values of the wax and diesel in testing data were unsatisfactory. In addition, Fig. 5 performed the relevant features of liquid compositions predictions and their contribution rankings. The top 3 most vital features related to wax production were the feed intake, reaction temperature, and vapour residue time, all about reactor conditions. The prediction features ranking for the diesel were particle size, vapour residue time, and feed intake. When predicting the middle carbon number products, the physical properties contributed more than the reactor conditions. For the low carbon number product, like gasoline, carbon, and hydrogen ratios were the most vital features for the product predictions. Likewise, the elemental ratios gave the same effects as gasoline on aromatics. This result implied that the reaction conditions decided the high carbon number products generation, such as wax. However, the predictions of specific chemicals, like aromatics, were heavily determined by the chemical properties of the feedstock. (Fig. 6 and 7).

Table 3 also depicted the R^2 and RMSE values of the machine learning results for gas compositions. The training R^2 of the model from the fitrensemble algorithm for C1, C2, C3, C4, and trace gas predictions were 0.981, 0.991, 0.983, 0.991, and 0.992, respectively. The prediction performance of testing data was much lower when compared to the training data. In common sense, the carbon numbers of fixed gas from pyrolysis were less than 4, such as CO, CO_2, and the low carbon number of alkanes and olefins. Except for the trace gas, fixed gas outputs mainly depended on setup temperatures in experimental research. The prediction models confirmed that the primary features for C1 to C4 generations were the pyrolysis temperatures. In terms of the predictions of C1 and C2 products, hydrogen content in feedstocks also played an important role. It might be because of the effects of the unsaturated linkages. However, for the predictions of C3 and C4 products, the top 3 features were all related to the reactor conditions, which is like the wax production. It showed that the reactor operating parameters were the primary factors of general pyrolysis predictions, like wax and gas. While for some other chemicals, such as aromatics and light oil, the feedstock properties were still the first determinant.

As discussed above, the prediction of other factors than the target composition and yield was significant to map the pyrolysis and extend the capacity of model-based process design. Despite the R^2 values of the composition prediction being about 0.99 for most training data, it is still unknown to clarify the black-box nature by quantifying the

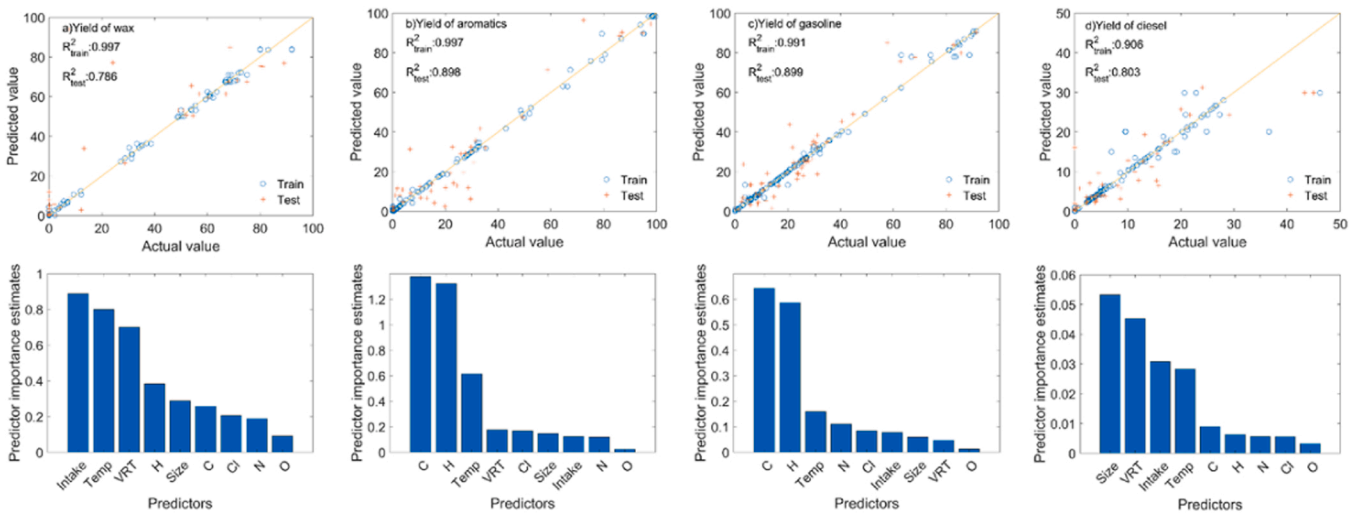


Fig. 6. Plots of the actual and predicted values obtained from the fitrensemble with Dataset 2 for liquid compositions: (a) yield of wax; (b) yield of aromatics, (c) yield of gasoline, and (d) yield of diesel.

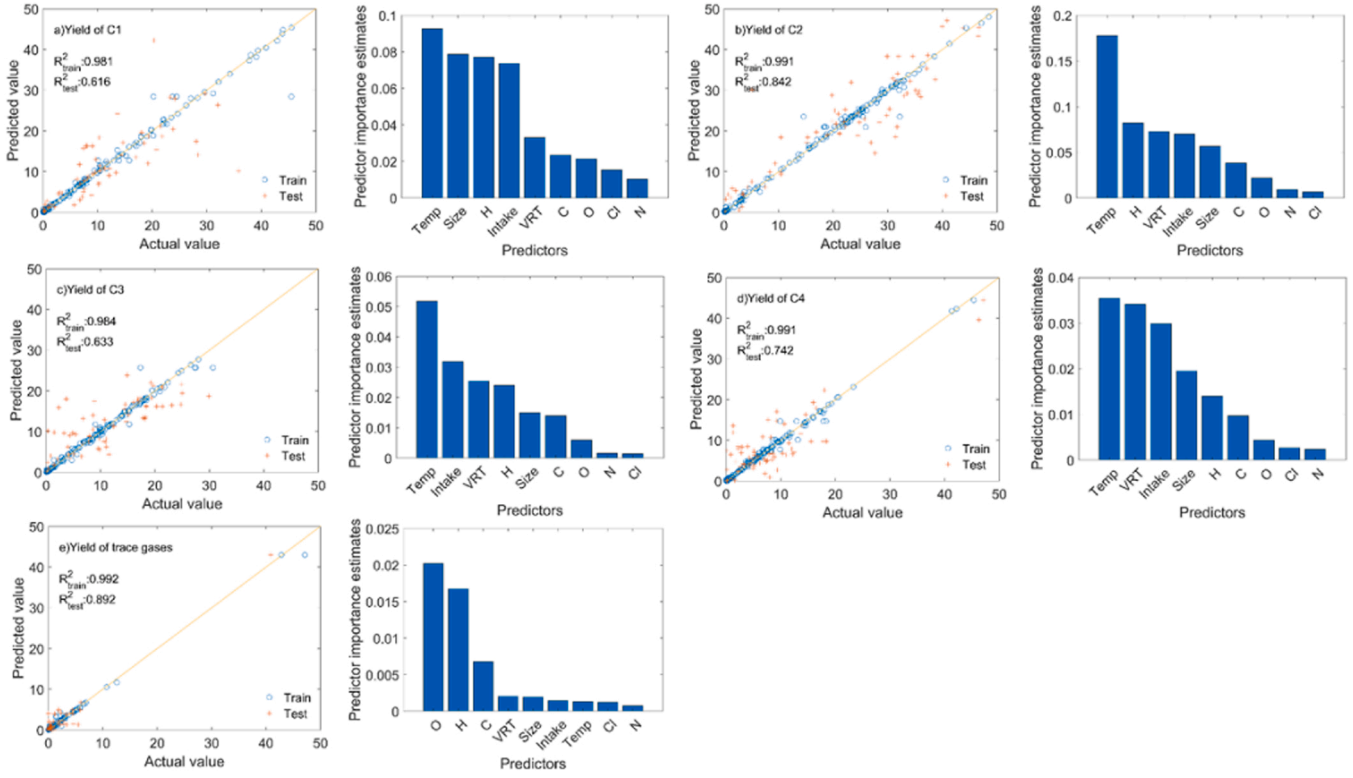


Fig. 7. Plots of the actual and predicted values obtained from the fitrensemble with Dataset 2 for gas compositions yields: (a) of C1; (b) of C2, (c) of C3, (d) of C4, and (d) of the trace gas.

contributions of input parameters. A problem previously pointed out was that the black-box nature limited data and model interpretability [34]. Waste recycling systems were complex systems spanning a range of areas. Hence, developing comprehensive ML models for waste management schemes requires experience across different disciplines. A review of ML methods in organic solid waste treatment discussed various thermochemical approaches [44]. The authors noticed that combining different ML techniques in integrated models or ML techniques with other innovative approaches was promising.

4. Conclusions

This work evaluated four ML algorithms to develop an applicable prediction model of plastic pyrolysis for valuable fuel and chemicals production. The fitrensemble mode of DT with ultimate analysis input carried out the outstanding capability of the products prediction. When using Dataset 1 to predict liquid yields, DT and ANN performed R^2 values of 0.984 and 0.926 for the training data, which were much higher than the 0.592 of GP and 0.583 of SVM. In the cases of mass balance and compositions prediction, the fitrensemble mode of DT also performed excellent results. The R^2 values for training data of solid, liquid, and gas

yields were above 0.99. Liquid compositions of wax, aromatics, and gasoline showed R^2 values of over 0.99 for the training data. Only diesel gave 0.906 of R^2 value. The gas compositions also gave no less than 0.98 R^2 values for the training data. Besides, investigation of the input feature importance and correlation for the feedstocks and the reactor conditions implied some connections existed between the input features, which might overcome the data missing by input incompletion. The operating condition (reaction temperature and residence time) and even heat transfer (particle size, feedstock intake) were the primary factors of general pyrolysis predictions, like solid, liquid, wax, and gas. But for the predictions of some chemicals, such as aromatics, styrene, and light oil, the elemental composition of the feedstock was still the first determinant.

CRediT authorship contribution statement

Yi Cheng: Investigation, Formal analysis, Writing – original draft. **Ecrin Ekici:** Formal analysis, Writing – review & editing. **Güray Yildiz:** Supervision, Writing – review & editing, Funding acquisition. **Yang Yang:** Writing – review & editing, Funding acquisition. **Brad Coward:** Writing – review & editing. **Jiawei Wang:** Supervision, Investigation, Writing – review & editing, Funding acquisition.

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.

Acknowledgements

The work was supported by an Institutional Links grant (No. 527641843), under the Turkey partnership. The grant is funded by the UK Department for Business, Energy, and Industrial Strategy together with the Scientific and Technological Research Council of Turkey (TÜBİTAK; Project no. 119N302) and delivered by the British Council. The author Yi Cheng and Jiawei Wang would like to acknowledge the Marie Skłodowska Curie Actions Fellowships by The European Research Executive Agency (H2020-MSCA-IF-2020, no. 101025906). The author Jiawei Wang would also like to acknowledge the support from Guangdong Science and Technology Program, No. 2021A0505030008.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jaap.2023.105857.

References

[1] D. Lazarevic, E. Aoustin, N. Buclet, N. Brandt, Plastic waste management in the context of a European recycling society: comparing results and uncertainties in a life cycle perspective, Resour. Conserv. Recycl. 55 (2) (2010) 246–259, https://doi.org/10.1016/j.resconrec.2010.09.014.
[2] H. Sardon, A.P. Dove, Plastics recycling with a difference, Sciences (80-.) 360 (6387) (2018) 380–381, https://doi.org/10.1126/science.aat4997.
[3] R. Geyer, J.R. Jambeck, K.L. Law, Production, use, and fate of all plastics ever made, Sci. Adv. 3 (7) (2017) 25–29, https://doi.org/10.1126/sciadv.1700782.
[4] D.C. Wilson, et al., 'Wasteaware' benchmark indicators for integrated sustainable waste management in cities, Waste Manag. 35 (2015) 329–342, https://doi.org/10.1016/j.wasman.2014.10.006.
[5] W.W.Y. Lau, et al., Evaluating scenarios toward zero plastic pollution, Sciences (80-) 369 (6509) (2020) 1455–1461, https://doi.org/10.1126/SCIENCE.ABA9475.
[6] M. Ilyas, W. Ahmad, H. Khan, S. Yousaf, K. Khan, S. Nazir, Plastic waste as a significant threat to environment - a systematic literature review, Rev. Environ. Health 33 (4) (2018) 383–406, https://doi.org/10.1515/reveh-2017-0035.
[7] E. Masanet, R. Auer, D. Tsuda, T. Barillot, A. Baynes, An assessment and prioritization of 'design for recycling' guidelines for plastic components, IEEE Int. Symp. Electron. Environ. (2002) 5–10, https://doi.org/10.1109/isee.2002.1003229.
[8] S.D. Anuar Sharuddin, F. Abnisa, W.M.A. Wan Daud, M.K. Aroua, A review on pyrolysis of plastic wastes, Energy Convers. Manag. 115 (2016) 308–326, https://doi.org/10.1016/j.enconman.2016.02.037.
[9] S.L. Wong, N. Ngadi, T.A.T. Abdullah, I.M. Inuwa, Current state and future prospects of plastic waste as source of fuel: a review, Renew. Sustain. Energy Rev. 50 (2015) 1167–1180, https://doi.org/10.1016/j.rser.2015.04.063.
[10] R. Khatun, H. Xiang, Y. Yang, J. Wang, G. Yildiz, Bibliometric analysis of research trends on the thermochemical conversion of plastics during 1990–2020, J. Clean. Prod. 317 (2021), 128373, https://doi.org/10.1016/j.jclepro.2021.128373.
[11] N. Li, H. Liu, Z. Cheng, B. Yan, G. Chen, S. Wang, Conversion of plastic waste into fuels: a critical review, J. Hazard. Mater. 424 (PB) (2022), 127460, https://doi.org/10.1016/j.jhazmat.2021.127460.
[12] M.S. Qureshi, et al., Pyrolysis of plastic waste: opportunities and challenges, J. Anal. Appl. Pyrolysis 152 (February) (2020), https://doi.org/10.1016/j.jaap.2020.104804.
[13] S. Armenise, et al., Plastic waste recycling via pyrolysis: a bibliometric survey and literature review, J. Anal. Appl. Pyrolysis 158 (2021), https://doi.org/10.1016/j.jaap.2021.105265.
[14] V.L. Mangesh, S. Padmanabhan, P. Tamizhdurai, A. Ramesh, Experimental investigation to identify the type of waste plastic pyrolysis oil suitable for conversion to diesel engine fuel, J. Clean. Prod. 246 (2020), 119066, https://doi.org/10.1016/j.jclepro.2019.119066.
[15] P. Das, P. Tiwari, Valorization of packaging plastic waste by slow pyrolysis, Resour. Conserv. Recycl. 128 (2017) (2018) 69–77, https://doi.org/10.1016/j.resconrec.2017.09.025.
[16] Y. Sakata, M.A. Uddin, A. Muto, Degradation of polyethylene and polypropylene into fuel oil by using solid acid and non-acid catalysts, J. Anal. Appl. Pyrolysis 51 (1) (1999) 135–155, https://doi.org/10.1016/S0165-2370(99)00013-3.
[17] A. Marcilla, M.I. Beltrán, R. Navarro, Thermal and catalytic pyrolysis of polyethylene over HZSM5 and HUSY zeolites in a batch reactor under dynamic conditions, Appl. Catal. B Environ. 86 (1–2) (2009) 78–86, https://doi.org/10.1016/j.apcatb.2008.07.026.
[18] Y. Liu, J. Qian, J. Wang, Pyrolysis of polystyrene waste in a fluidized-bed reactor to obtain styrene monomer and gasoline fraction, Fuel Process. Technol. 63 (1) (2000) 45–55, https://doi.org/10.1016/S0378-3820(99)00066-1.
[19] I. Ahmad, et al., Pyrolysis study of polypropylene and polyethylene into premium oil products, Int. J. Green Energy 12 (7) (2015) 663–671, https://doi.org/10.1080/15435075.2014.880146.
[20] R. Miranda, J. Yang, C. Roy, C. Vasile, Vacuum pyrolysis of commingled plastics containing PVC. I. Kinetic study, Polym. Degrad. Stab. 72 (3) (2001) 469–491, https://doi.org/10.1016/S0141-3910(01)00048-9.
[21] I. Dubdub, M. Al-Yaari, Pyrolysis of mixed plastic waste: II. Artificial neural networks prediction and sensitivity analysis, Appl. Sci. 11 (18) (2021), https://doi.org/10.3390/app11188456.
[22] S. Ascher, I. Watson, S. You, Machine learning methods for modelling the gasification and pyrolysis of biomass and waste, Renew. Sustain. Energy Rev. (2021), 111902.
[23] T.Y. Li, H. Xiang, Y. Yang, J. Wang, G. Yildiz, Prediction of char production from slow pyrolysis of lignocellulosic biomass using multiple nonlinear regression and artificial neural network, J. Anal. Appl. Pyrolysis 159 (July) (2021), 105286, https://doi.org/10.1016/j.jaap.2021.105286.
[24] H. Wei, K. Luo, J. Xing, J. Fan, Predicting co-pyrolysis of coal and biomass using machine learning approaches, Fuel 310 (PA) (2022), 122248, https://doi.org/10.1016/j.fuel.2021.122248.
[25] H. Yaka, M. Akin, O. Yucel, H. Sadikoglu, A comparison of machine learning algorithms for estimation of higher heating values of biomass and fossil fuels from ultimate analysis, Fuel 320 (August 2021) (2022), 123971, https://doi.org/10.1016/j.fuel.2022.123971.
[26] A.Y. Mutlu, O. Yucel, An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification, Energy 165 (2018) 895–901, https://doi.org/10.1016/j.energy.2018.09.131.
[27] D. Baruah, D.C. Baruah, Modeling of biomass gasification: a review, Renew. Sustain. Energy Rev. 39 (2014) 806–815, https://doi.org/10.1016/j.rser.2014.07.129.
[28] F. Cheng, H. Luo, L.M. Colosi, Slow pyrolysis as a platform for negative emissions technology: an integration of machine learning models, life cycle assessment, and economic analysis, Energy Convers. Manag. 223 (August) (2020), 113258, https://doi.org/10.1016/j.enconman.2020.113258.
[29] H. Cao, Y. Xin, Q. Yuan, Prediction of biochar yield from cattle manure pyrolysis via least squares support vector machine intelligent approach, Bioresour. Technol. 202 (2016) 158–164, https://doi.org/10.1016/j.biortech.2015.12.024.
[30] N.E.I. Karabadji, H. Seridi, F. Bousetouane, W. Dhifli, S. Aridhi, An evolutionary scheme for decision tree construction, Knowl. -Based Syst. 119 (2017) 166–177, https://doi.org/10.1016/j.knosys.2016.12.011.
[31] T. Tsoka, X. Ye, Y.Q. Chen, D. Gong, X. Xia, Explainable artificial intelligence for building energy performance certificate labelling classification, J. Clean. Prod. 355 (March) (2022), 131626, https://doi.org/10.1016/j.jclepro.2022.131626.
[32] A. Altikat, M.H. Alma, Prediction carbonization yields and the sensitivity analyses using deep learning neural networks and support vector machines, Int. J. Environ. Sci. Technol. (Demirbas 2001) (2022), https://doi.org/10.1007/s13762-022-04407-1.

[33] E. Yapıcı, H. Akgün, K. Özkan, Z. Günkaya, A. Özkan, M. Banar, Prediction of gas product yield from packaging waste pyrolysis: support vector and Gaussian process regression models, Int. J. Environ. Sci. Technol. (0123456789) (2022), https://doi.org/10.1007/s13762-022-04013-1.

[34] T. Hastie, J. Friedman, R. Tibshirani, Neural Networks. The Elements of Statistical Learning. Springer Series in Statistics, Springer New York, New York, NY, 2001, pp. 347–369, https://doi.org/10.1007/978-0-387-21606-5_11.

[35] F. Elmaz, Ö. Yücel, A.Y. Mutlu, Predictive modeling of biomass gasification with machine learning-based regression methods, Energy 191 (2020), 116541.

[36] B.R. Hough, D.A.C. Beck, D.T. Schwartz, J. Pfaendtner, Application of machine learning to pyrolysis reaction networks: reducing model solution time to enable process optimization, Comput. Chem. Eng. 104 (2017) 56–63, https://doi.org/10.1016/j.compchemeng.2017.04.012.

[37] S. Zinchik, et al., Accurate characterization of mixed plastic waste using machine learning and fast infrared spectroscopy, ACS Sustain. Chem. Eng. 9 (42) (2021) 14143–14151, https://doi.org/10.1021/acssuschemeng.1c04281.

[38] B. Kunwar, H.N. Cheng, S.R. Chandrashekaran, B.K. Sharma, Plastics to fuel: a review, Renew. Sustain. Energy Rev. 54 (2016) 421–428, https://doi.org/10.1016/j.rser.2015.10.015.

[39] Q. Tang, et al., Prediction of bio-oil yield and hydrogen contents based on machine learning method: effect of biomass compositions and pyrolysis conditions, Energy Fuels 34 (9) (2020) 11050–11060, https://doi.org/10.1021/acs.energyfuels.0c01893.

[40] X. Zhu, Y. Li, X. Wang, Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions, Bioresour. Technol. 288 (May) (2019), 121527, https://doi.org/10.1016/j.biortech.2019.121527.

[41] A. Sarwar, et al., Synthesis and characterization of biomass-derived surface-modified activated carbon for enhanced CO_2 adsorption, J. CO_2 Util. 46 (January) (2021), 101476, https://doi.org/10.1016/j.jcou.2021.101476.

[42] S.R. Naqvi, Y. Uemura, S. Yusup, Y. Sugiura, N. Nishiyama, In situ catalytic fast pyrolysis of paddy husk pyrolysis vapors over MCM-22 and ITQ-2 zeolites, J. Anal. Appl. Pyrolysis 114 (2015) 32–39, https://doi.org/10.1016/j.jaap.2015.04.003.

[43] H.W. Ryu, D.H. Kim, J. Jae, S.S. Lam, E.D. Park, Y.K. Park, Recent advances in catalytic co-pyrolysis of biomass and plastic waste for the production of petroleum-like hydrocarbons, Bioresour. Technol. 310 (February) (2020), 123473, https://doi.org/10.1016/j.biortech.2020.123473.

[44] H. nan Guo, S. biao Wu, Y. jie Tian, J. Zhang, H. tao Liu, Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: a review, Bioresour. Technol. 319 (September 2020) (2021), 124114, https://doi.org/10.1016/j.biortech.2020.124114.