

**MACHINE LEARNING MODELS FOR MUNICIPAL SOLID
WASTE PYROLYSIS**

BY

Temirlan Shakhmanov

A Thesis Presented to the

COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

Chemical Engineering

May 2025

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES

This thesis, written by **Temirlan Shakhmanov** under the direction of his thesis advisor and approved by his/her thesis committee, has been presented and accepted by the Dean, College of Graduate and Interdisciplinary Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN CHEMICAL ENGINEERING**



Dr. Saad Fahaid Khalaf Al-Afnan
Department Chairman



Dr. Suliman S. Al-Homidan
Dean, College of Graduate and
Interdisciplinary Studies

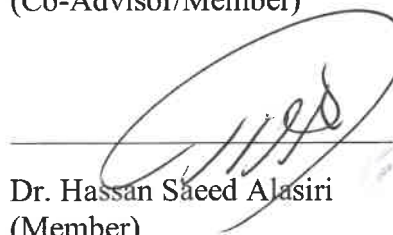
Date



Dr. Mohammed Fahad Al-Khater
(Advisor)



Dr. Omar Abdelaziz
(Co-Advisor/Member)



Dr. Hassan Saeed Alasiri
(Member)



© Temirlan Shakhmanov

2025

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to:

- King Fahd University of Petroleum and Minerals and the Chemical Engineering Department
- Dr. Mohammed Fahad Al-Khater, Dr. Omar Abdelaziz and Dr. Hassan Saeed Alasiri
- My family
- My friends and colleagues

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	IV
TABLE OF CONTENTS.....	V
LIST OF TABLES.....	VIII
LIST OF FIGURES.....	IX
LIST OF ABBREVIATIONS.....	X
ABSTRACT.....	XII
ملخص الرسالة.....	XIII
CHAPTER 1 INTRODUCTION.....	1
1.1 Background of the Study.....	1
1.2 Problem Statement.....	9
1.3 Objectives of the Study.....	10
1.4 Research Questions.....	11
1.5 Significance of the Study.....	12
1.6 Scope and Limitations.....	14
1.7 Organization of the Thesis.....	15
CHAPTER 2 LITERATURE REVIEW.....	17
2.1 Background and Motivation.....	17
2.2 Evolution of Pyrolysis for Plastic Waste Management.....	18
2.3 Key Factors Influencing Pyrolysis Yield.....	19

2.4	Machine Learning in Pyrolysis Modeling.....	22
2.5	Previous Studies on ML-Based Yield Prediction.....	24
2.6	Identified Research Gaps.....	26
2.7	Summary.....	28
CHAPTER 3 METHODOLOGY.....		30
3.1	Data Collection and Preprocessing.....	30
3.2	Feature Engineering.....	32
3.3	Model Selection and Justification.....	33
3.4	Hyperparameter Tuning.....	35
3.5	Evaluation Metrics.....	37
3.6	Feature Importance Analysis.....	39
3.7	Summary of Methodology.....	40
CHAPTER 4 RESULTS AND DISCUSSION.....		42
4.1	Overview.....	42
4.2	Machine Learning Model Performance.....	42
4.3	Feature Importance and SHAP Analyses.....	44
4.4	Symbolic Regression Results	47
4.5	Discussion.....	49
4.6	Summary of Chapter 4.....	50
CHAPTER 5 CONCLUSION.....		53
5.1	Overview.....	53
5.2	Key Findings.....	54
5.3	Engineering Implications.....	58
5.4	Future Work.....	60
REFERENCES.....		63

VITAE.....	73
------------	----

LIST OF TABLES

Table 1	Comparison of Waste Disposal Methods	1
Table 2	Comparison of Pyrolysis Types	4
Table 3	Research Objectives and Expected Outcomes	11
Table 4	Scope vs Limitations of the Study	14
Table 5	Key Parameters Affecting Pyrolysis Outcomes	20
Table 6	Common ML Algorithms Used in Pyrolysis Studies	23
Table 7	Statistical Summary of Input Features	32
Table 8	Machine Learning Models Overview	34
Table 9	Evaluation Metrics	38
Table 10	Machine Learning Model Performance Metrics	43

LIST OF FIGURES

Figure 1	Flow Diagram of Plastic Waste Fate and Recovery Paths	2
Figure 2	ML Applications in Pyrolysis Modeling	6
Figure 3	ML-Pyrolysis Framework Aligned with Vision 2030	13
Figure 4	ML Workflow for Pyrolysis Modeling	22
Figure 5	Optimal Hyperparameter Settings	37
Figure 6	Comparative Feature Importance across Tree-Based Models	46
Figure 7	SHAP analysis	47
Figure 8	Symbolic Expression using Feyn QLatice	48

LIST OF ABBREVIATIONS

AdaBoost	:	Adaptive Boosting
AI	:	Artificial Intelligence
ANN	:	Artificial Neural Network
CatBoost	:	Categorical Boosting
CNNs	:	Cnvolutional Neural Networks
COVID--19	:	COronaVirus Disease of 2019
CFD	:	Computational Fluid Dynamics
DBSCAN	:	Density-Based Spatial Clustering of Applications with Noise
EPR	:	Extended Producer Responsibility
FCC	:	Fluid catalytic cracking
GBDT	:	Gradient Boosting Decision Trees
HDPE	:	High Density Poly Ethylene
IQR	:	Interquartile Range
LCA	:	Life Cycle Analysis
LDPE	:	Low Density Poly Ethylene
LightGBM	:	Light Gradient-Boosting Machine

MAE	:	Mean Absolute Error
MAP	:	Microwave-Assisted Pyrolysis
ML	:	Machine Learning
MSW	:	Municipal Solid Waste
PET	:	Polyethylene terephthalate
PP	:	Polypropylene
PS	:	Polystyrene
PVC	:	Polyvinyl chloride
RF	:	Random Forest
QLattice	:	Quantum Lattice
RMSE	:	Root Mean Square Error
RNNs	:	Recurrent Neural Networks
ROI	:	Return On Investment
SHAP	:	SHapley Additive exPlanations
SVR	:	Support Vector Machines
TEA	:	Techno-Economic Analysis
XGBoost	:	Extreme Gradient Boosting

ABSTRACT

Full Name : Temirlan Shakhmanov

Thesis Title : MACHINE LEARNING MODELS FOR MUNICIPAL SOLID WASTE PYROLYSIS

Major Field : Chemical Engineering

Date of Degree : May,2025

The rising complexity of municipal solid waste (MSW) and aggravated by expanding plastic usage necessitates new and sustainable MSW management alternatives. Pyrolysis, a thermal conversion technology for the transformation of MSW to energy-rich fuels, is a waste-to-energy alternative. Yet, accurate prediction of pyrolysis liquid product is a big challenge owing to the nonlinearity involved in the diversity of feedstocks and operating conditions. The conventional kinetic models are incapable of this richness with the fidelity required. In this study, the application of sophisticated machine learning (ML) methods to simulate the pyrolysis process and predict the liquid yield from a blend of municipal solid waste (MSW) and plastic waste is explored.

A suitably chosen dataset of over 300 experimental points was collected from the literature, and the detailed list of input variables consisted of elemental composition (C, H, O, N, Cl), polymeric content (HDPE, LDPE, PP, PET, PVC, PS), particle size, heating rate, and temperature. Feature engineering and normalization techniques were applied for preprocessing the dataset, and training and testing of seven machine learning models like Random Forest, Gradient Boosting Decision Trees (GBDT), XGBoost, CatBoost, LightGBM, AdaBoost-SVR, and Feyn symbolic regression were done as a result. R^2 , RMSE, MAE.

Among the models, Random Forest and GBDT had the top test R^2 values (0.8952 and 0.8940, respectively). Feyn symbolic regression, though less accurate (test $R^2 = 0.827$), gave readable algebraic equations that reflected key relationships between variables. Feature importance comparison across models identified temperature, particle size, and polymeric fractions (PP, HDPE, LDPE, PET) as the highest contributors to the yield variation. SHAP analysis provided an additional dimension of model interpretability so that input-output interaction insights were meaningful.

This research provides a robust, data-driven methodology that integrates high-performance ensemble modeling with understandable symbolic modeling. The outcomes support the development of smart and adaptive pyrolysis systems and open doors to sustainable waste management and facilitate national initiatives like Saudi Vision 2030.

ملخص الرسالة

الاسم الكامل: تيمير لان شاهمانوف

عنوان الرسالة: نماذج التعلم الآلي لتحلل النفايات الصلبة البلدية

التخصص: الهندسة الكيميائية

تاريخ الدرجة العلمية: مايو 2025

تزداد تعقيدات النفايات الصلبة البلدية بشكل مستمر، ويتفاقم الوضع بسبب الزيادة في استخدام البلاستيك، مما يستدعي إيجاد بدائل جديدة ومستدامة لإدارة هذه النفايات. يُعد التحلل الحراري إحدى تقنيات التحويل الحراري التي تُحوّل النفايات الصلبة إلى وقود غني بالطاقة، وهو خيار واعد ضمن استراتيجيات تحويل النفايات إلى طاقة. ومع ذلك، فإن التنبؤ الدقيق بكمية السائل الناتج عن التحلل يُعد تحديًا كبيرًا بسبب الطبيعة غير الخطية لتنوع المواد الأولية وظروف التشغيل المختلفة. فالنماذج الحركية التقليدية غير قادرة على محاكاة هذا التنوع بالدقة المطلوبة. في هذه الدراسة، تم استكشاف تطبيق نماذج تعلم آلي متقدمة لمحاكاة عملية التحلل الحراري والتنبؤ بكمية السائل الناتج من مزيج من النفايات الصلبة البلدية والنفايات البلاستيكية.

تم جمع مجموعة بيانات مناسبة تتضمن أكثر من 300 نقطة تجريبية من الأدبيات العلمية، وتضمنت قائمة المتغيرات المدخلة التركيب العنصري (الكربون، الهيدروجين، الأوكسجين، النيتروجين، الكلور)، ومحتوى البوليمرات (البولي إيثيلين عالي الكثافة، البولي إيثيلين منخفض الكثافة، البولي بروبيلين، البولي إيثيلين تيرفتالات، البولي فينيل كلورايد، البولي ستايرين)، وحجم الجسيمات، ومعدل التسخين، ودرجة الحرارة. وقد تم تطبيق تقنيات هندسة الميزات والتطبيع كجزء من معالجة البيانات الأولية، ثم تم تدريب واختبار سبعة نماذج تعلم آلي، وهي: الغابة العشوائية، أشجار اتخاذ القرار المعززة تدريجيًا، المعزز شديد التدرج، المعزز باستخدام كات، المعزز الخفيف، معزز آدا - الانحدار الداعم، والانحدار الرمزي فين.

أظهرت نتائج نموذجي الغابة العشوائية وأشجار اتخاذ القرار المعززة تدريجيًا أعلى قيم معامل التحديد (0.8952 و 0.8940 على التوالي). (أما الانحدار الرمزي فين، فرغم أنه كان أقل دقة بمعامل التحديد 0.827)، إلا أنه قدم معادلات جبرية قابلة للفهم تعكس العلاقات الأساسية بين المتغيرات. كما أظهرت مقارنة أهمية الميزات عبر النماذج أن درجة الحرارة، حجم الجسيمات، ونسب البوليمرات (البولي بروبيلين، البولي إيثيلين عالي الكثافة، البولي إيثيلين منخفض الكثافة، البولي إيثيلين تيرفتالات (كانت من بين أهم العوامل المؤثرة في تغير العائد السائل. وقد أضاف تحليل بُعدًا إضافيًا لتفسير النماذج، مما وفر رؤى ذات معنى حول تفاعلات المدخلات (SHAP) القيم التوضيحية التراكمية والمخرجات.

تقدم هذه الدراسة منهجية قوية قائمة على البيانات، تجمع بين النمذجة التجميعية عالية الأداء والنمذجة الرمزية القابلة للتفسير. وتدعم هذه النتائج تطوير أنظمة تحلل حراري ذكية وقابلة للتكيف، كما تفتح آفاقًا جديدة نحو إدارة مستدامة للنفايات، وتساهم في تحقيق مبادرات وطنية مثل رؤية السعودية 2030.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Municipal Solid Waste (MSW) management has become an acute global issue as a result of fast urbanization, industrialization, and shifting consumption patterns. This accelerating waste generation has placed unsustainable pressures on conventional disposal practices, with landfilling and incineration continuing to be major waste management practices despite their considerable environmental disadvantages (**Table 1** for a comparative view) (Daskalopoulos et al., 1998). The COVID-19 pandemic has further increased the waste crisis, particularly by increased use of single-use plastics and biomedical waste (Leal Filho et al., 2021). Under such a context, waste-to-energy technologies, particularly pyrolysis, offer an attractive pathway for addressing such problems by transforming MSW and plastic waste into valuable fuels and chemicals (Z. Wang et al., 2021).

Table 1 Comparison of Waste Disposal Methods (Daskalopoulos et al., 1998)

Method	Landfilling	Incineration	Pyrolysis
Advantages	Low cost, established	Volume reduction, energy recovery	Converts waste to fuel, scalable
Disadvantages	Space-intensive, leachate, methane emissions	Air pollution, toxic ash, high capital cost	High setup cost, emissions control required
Environmental Impact	High negative	High negative (if uncontrolled)	Medium to low (if optimized)

Moreover, growing heterogeneity of plastics poses problems for recycling and requires the development of other recovery mechanisms; this development is reflected in **Figure 1**, which outlines the potential directions and recovery pathways of plastic waste (Schmidt & Laner, 2021). Traditional mechanical recycling processes are usually troubled by contamination, degradation of polymer properties, and economic infeasibility in recycling mixed or contaminated plastics streams (Khalid et al., 2022). Chemical recycling incorporating pyrolysis has been suggested as an attractive alternative that enables breakdown of complicated polymers into less complex chemical species and thereby provides an opportunity for high-value material recycling (Peng et al., 2022).

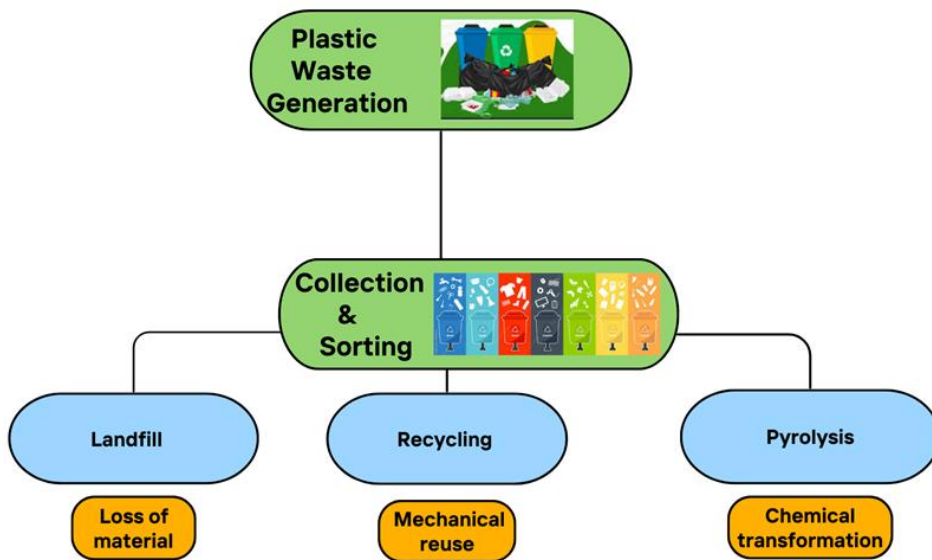


Figure 1 Flow Diagram of Plastic Waste Fate and Recovery Paths (Schmidt & Laner, 2021)

Nations have reacted to legislative and environmental drivers by launching circular economy strategies for plastic waste management and resource recovery (Yamamoto & Eva, 2022). International policies on extended producer responsibility (EPR), plastic bans, and investments in recycling infrastructure have been launched, though facing challenges

to scale up sustainable technologies (Cascone & Gagliano, 2022). Scientifically, there is increased focus on designing integrated waste management systems with interdependent environmental, economic, and technological factors for the promotion of long-term sustainability (Paul et al., 2018).

Within this broader push towards circularity and sustainable development, pyrolysis is one solution that is particularly apt to convert waste plastics into recyclable energy carriers and feedstock chemicals. Both as a mitigation and valorization strategy, pyrolysis can help decarbonize and diversify energy while minimizing the environmental footprint of persistent plastic pollution (Areeprasert et al., 2017).

Pyrolysis is a thermal-induced chemical conversion of organic materials under oxygen-free conditions. Pyrolysis offers possibilities to produce plastic and biomass wastes as marketable products such as bio-oil, syngas, and char. Various pyrolysis processes have been developed in an attempt to achieve maximum product yield and energy efficiency.

A comparative summary of these processes with advantages and typical outputs is presented in **Table 2** (Lam, 2016). However, the yield and pyrolysis product quality are influenced by numerous interacting variables including temperature, heating rate, feedstock composition, and catalyst type (Kajiwara et al., 2021).

Table 2 Comparison of Pyrolysis Types (Lam, 2016)

Type	Catalytic Pyrolysis	Flash Pyrolysis	Microwave-assisted	Co-pyrolysis
Heating Source	Conventional furnace	Rapid heating	EM radiation	Furnace/combined
Temp Range (°C)	400–600	>500	300–600	400–700
Key Advantages	Selective output, lower temperature	Fast conversion	Uniform heating, fast startup)	Synergistic yields, less char
Products	Liquid fuels, syngas	Bio-oil, light gases	Liquid fuels, light gases	Enhanced oil + gas

Catalytic pyrolysis has been extensively studied due to its capability of reducing reaction temperatures, improving product selectivity, and reducing tar formation (Aisien et al., 2021). Co-pyrolysis, or the simultaneous pyrolysis of plastics with biomass or other organic materials, has been shown to exhibit synergistic effects that improve overall thermal efficiency and yield distributions (Hu et al., 2020). Microwave pyrolysis, where electromagnetic radiation is used as the source of heat, has benefits of quick heating, energy savings, and control of the process (Putra et al., 2022).

Flash pyrolysis and high-temperature processes, especially in fluidized bed reactors, have been employed for fast waste conversion to liquid fuels and value-added chemicals (KLAIMY et al., 2021). Design improvements, such as downdraft, bubbling fluidized bed, and rotary kiln designs, have also increased the conversion efficiencies and scale-up (Tokmurzin et al., 2022). Nonetheless, the built-in complexity in feedstocks of waste like

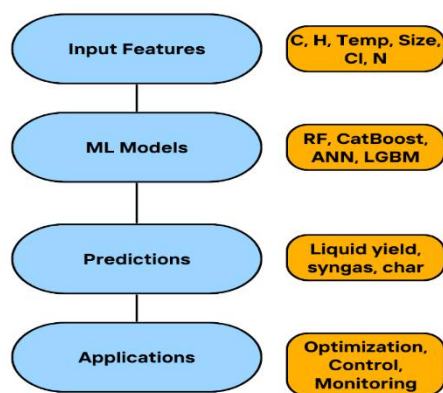
mixed plastic streams, impurities, and uneven moisture and ash content makes consistent product quality and conversion rates challenging to achieve (Martínez-Narro et al., 2022). Moreover, environmental performance within pyrolysis processes is also dependent on process conditions and feedstock properties. Based on life cycle assessments, optimally, pyrolysis has better performance relative to mechanical recycling and incineration in energy recovery as well as greenhouse gas emissions (Jeswani et al., 2021). Emissions and residues of hazardous pollutants in processes also need to be tightly regulated to conform to environmental regulations (Fu et al., 2022). The combination of renewable energy sources like microwaves or solar-assisted heating with carbon capture technologies further increases the scope of pyrolysis in sustainable waste valorization (Cudjoe & Wang, 2022). Despite enhanced pyrolysis technology, one of the biggest challenges remains forecasting product yields from heterogeneous feedstocks typical of MSW. Kinetic and mechanistic models are generally poor in modeling the nonlinear dynamics and variability in real plastic waste streams (Liu et al., 2022). To overcome this limitation, data-driven approaches have been proposed, and machine learning (ML) techniques have been found extremely promising in modeling complex systems (Shufian et al., 2022).

The predictive capability of ML algorithms stems from the ability to learn complex patterns in data without being programmed with physical laws or assumptions. This flexibility allows them to outperform traditional statistical or mechanistic models when one is handling big, nonlinear, and high-dimensional data (Raj et al., 2023). Moreover, ML models can be recalibrated and updated whenever new data emerges, making them adaptive solutions for real-time decision-making under dynamic processing conditions. These characteristics are particularly valuable for pyrolysis applications, for which feed

composition, operating conditions, and reactor configurations can be quite different (Veksha et al., 2020).

A number of studies have used machine learning to model pyrolysis with encouraging results. The overall process of such ML applications, from data input to predictions based on models and control of the process, is shown in **Figure 2**. Artificial neural networks (ANNs) and ensemble learning algorithms, for example, have been reported to perform better than other approaches in predicting product yields and optimizing process conditions (Ortiz et al., 2022). These models have also been employed to perform sensitivity analyses, and the significant drivers of variability in pyrolysis yields have been highlighted. With the use of feature selection methods and interpretable AI software like SHAP, scientists are now able to not only improve the performance of models but also provide meaningful explanations regarding the physical processes underlying the trends observed (Chow et al., 2017).

Figure 2 ML Applications in Pyrolysis Modeling (Shufian et al., 2022)



Machine learning algorithms have increasingly been applied to predict product yields from pyrolysis processes, e.g., the yield of pyro-oil, syngas, and char (Kajiwara et al., 2021). Advanced ML algorithms, including Random Forest, LightGBM, CatBoost, and neural

networks, have shown excellent prediction performance in different waste valorization applications (Hasanzadeh et al., 2022). Moreover, explainable ML techniques such as SHAP (Shapley Additive Explanations) are employed increasingly to describe model predictions and provide transparency concerning feature importance (Chow et al., 2018). Studies such as those by Al-Yaari and Dubdub have applied artificial neural networks to predict catalytic pyrolysis behavior of HDPE and mixed plastic waste (Areeprasert et al., 2017). Several researchers, including Abnisa et al., utilized feedforward neural networks along with backpropagation algorithms, i.e., Levenberg-Marquardt, for predicting plastic pyrolysis liquid fuel yield (Daskalopoulos et al., 1998). Further research has included the use of ensemble models that did a better job when dealing with noisy and heterogeneous data from experimental pyrolysis results (Hendrawati et al., 2023).

Aside from their forecast applications, machine learning algorithms are increasingly being applied towards process optimization. As an example, ML has been employed for the calculation of optimum temperature and catalyst proportions to maximize fuel production or minimize pollutant formation (Karimi Estahbanati et al., 2021). In some other instances, ML has been combined with decision-support programs such as multi-objective optimization procedures or utilized in conjunction with CFD models to produce a hybrid algorithm to study a system (Grimaldos-Osorio et al., 2023). These integrative approaches are a growing front in the application of ML to waste-to-energy systems, enabling more efficient design and control of pyrolysis operations (Cudjoe & Wang, 2022).

As ML methods become increasingly mainstream and robust, their ability to manage new forms of data—e.g., high-rate sensor data, infrared spectra, or gas chromatogram output—adds value to pyrolysis science and industry. A natural next step enabled by these advances

is evolution from predictive modeling to intelligent control and real-time monitoring of pyrolysis reactors. Thus, the use of ML in pyrolysis is growing from being an analytical tool to a full-fledged platform for process innovation and automation (Chen et al., 2021). Current studies have employed ML to predict yields in co-pyrolysis of plastics and biomass (Veksha et al., 2020). For example, Alabdrabalnabi et al. employed ML to predict pyro-char and pyro-oil yields for biomass-plastic mixtures, and Li et al. employed ML to optimize zeolite-catalytic pyrolysis (Hendrawati et al., 2023). Implementation of ML into waste-to-energy processes ensures optimum utilization efficiency, improves decision-making, and aligns to sustainable development (Yamamoto & Eva, 2022).

A growing body of literature also depicts the use of hybrid methods that combine ML with LCA, process simulation, or techno-economic analysis for evaluating the broader implications of plastic pyrolysis (Jeswani et al., 2021). These techniques are used to quantify the environmental, economic, and social benefits of using ML-optimized pyrolysis processes compared to conventional ones. Moreover, ML has been used in reactor-scale modeling to forecast emissions, byproduct yields, and conversion efficiencies as a function of operating conditions (J. Wang et al., 2021).

The scalability of ML platforms is also enhanced through the assistance of open-access datasets, cloud computing, and IoT device integration for real-time monitoring (Risco et al., 2021). With the growing application of sensor networks and smart manufacturing systems, ML-based pyrolysis reactors can be a core component of future waste valorization infrastructure. These trends validate the relevance of ML both as a method of data analysis but also as a primary support to convert to a circular, resource-efficient, and decarbonised economy (Siddique et al., 2022).

1.2 Problem Statement

Despite its advantages, pyrolysis is encumbered by the lack of realistic models to compute the partition of liquid, gaseous, and solid products, especially in application to complex and varying feedstocks like MSW. Reaction kinetics or thermodynamic-based models are limited in their assumptions and simplifications. Also, the application of ML to this area, as encouraging as it is, remains in an extremely nascent stage and is often plagued by shallow data, lack of explainability of results, and generalization performance against a wide range of feedstock types and pyrolysis plants (Moo et al., 2019). There exists a critical requirement for the design of robust, interpretable, and scalable ML models tailored towards predicting pyrolysis yield over a gigantic range of input conditions.

Although numerous ML-based papers have shown effective prediction performance with historical or test data, few address key issues like feature bias, model overfitting, or a lack of capability to extrapolate outside the training domain. This restricts them to use in industrial settings, where actual-time process variation and new waste streams are common (Grimaldos-Osorio et al., 2023). In addition, some published models represent the pyrolysis process as a black box with little explanation of the causal relationships between input and output variables. This lack of transparency in models creates problems for stakeholders who need validation, interpretation, or regulation of the system in use (Ortiz et al., 2022).

Another major limitation is heterogeneity and variability of training data for ML models. Experimental condition variability, measurement procedures, and feedstock characterization variability across studies render data pooling and model verification problematic. Absence of a common approach to data collection and preprocessing would

lead even the most advanced ML algorithms to generate suboptimal or misleading results (Chow et al., 2018). Thus, standardized data structures and open-access databases for reproducible and generalizable model development within the pyrolysis community are in immediate demand (Risco et al., 2021).

Furthermore, there have been few efforts to combine ML with emerging sensing and automation technologies for enabling real-time process control. Coupling of ML with online monitoring equipment would be able to offer predictive maintenance, yield prediction, as well as emissions monitoring, in essence rendering pyrolysis systems intelligent and adaptive platforms (Chen et al., 2021). Such coupling would offer closed-loop optimization and would also be compliant with Industry 4.0 principles, opening up possibilities for smart waste valorization systems (Siddique et al., 2022).

1.3 Objectives of the Study

The primary objective of this thesis is to develop and test advanced machine learning models for the prediction of liquid yield from pyrolysis of MSW and plastic waste. The objective of this research is to develop a predictive model that is not just accurate but also interpretable and capable of handling the complexity and variability of heterogeneous waste streams. Special emphasis is placed on the integration of feature engineering techniques and data normalization techniques for enhancing model performance, reliability, and scalability.

The key research questions, corresponding methodologies, and anticipated findings are tabulated in **Table 3** for easy reference.

Table 3 Research Objectives and Expected Outcomes (Raj et al., 2023)

Objective	Method/Action	Expected Outcome
Develop ML models	Train RF, LGBM, CatBoost	High accuracy predictions
Evaluate performance	Use R^2 , RMSE, MAE	Model comparison
Apply SHAP analysis	Feature interpretation	Increased transparency
Introduce FEYN framework	Feature normalization + yield scaling	Generalizable model across setups

1.4 Research Questions

This study aims to answer the following questions:

- What machine learning algorithm produces the most accurate forecasts of liquid yield in MSW pyrolysis?
- What are the most significant input features influencing the pyrolysis output?
- How are different ML models modeling the impact of feedstock and process parameters?
- Can explainable ML techniques enhance trust and understanding in predictive modeling?

These questions are asked not only to find the best-performing model but also to check the interpretability and applicability of such models in actual applications. By exhaustive

analysis of model performance and explainability, the study aims to shed light on the feasibility of ML for pyrolysis modeling. Additionally, insights into the ways different features affect yield predictions will facilitate the creation of smart data collection strategies and more efficient process optimization frameworks (Daskalopoulos et al., 1998).

The answer to these questions will inform the selection and deployment of predictive algorithms in pyrolysis systems and can ultimately contribute to the development of intelligent, adaptive waste conversion facilities. The focus on explainability is especially important to achieve transparency, which is often a limiting factor in the application of AI models in engineering practice (Ortiz et al., 2022).

1.5 Significance of the Study

This study provides a significant contribution to environmental engineering, waste management, and interdisciplinary data science research through the evidence of the potential of machine learning application in the optimization of processes of thermochemical waste conversion. The combination of ML models within pyrolysis systems is capable of changing the modeling and control of product yields so that municipal solid waste use is efficient and less dependent on landfills and incineration (Martínez-Narro et al., 2022).

This research also aligns with national goals under Saudi Vision 2030. The combination of ML with waste valorization for Vision 2030 pillars support is depicted in **Figure 3**. With the support of the innovation of waste-to-energy technologies, the research supports

national efforts to counteract environmental pollution, enhance resource recovery, and provide clean energy production (Yamamoto & Eva, 2022).

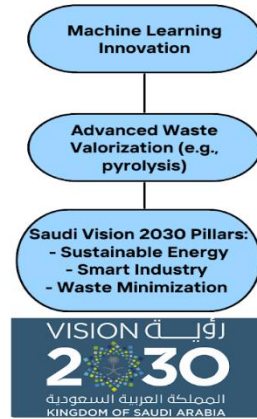


Figure 3 ML-Pyrolysis Framework Aligned with Vision 2030 (Yamamoto & Eva, 2022)

This is also well in line with the objectives of Vision 2030 to increase innovation, boost industrial productivity, and encourage the adoption of new technologies in priority sectors, particularly with data-inspired intelligent models for MSW pyrolysis. Additionally, the woes of plastic wastes are addressed while enabling local solutions on energy recovery; thus, it complements the environmental goals in Vision 2030, such as landfill reliance and enhancing municipal recycling rates (Meng et al., 2021).

Additionally, the study provides valuable advice to policymakers, engineers, and industry actors who seek to implement AI-based systems into environmental infrastructure. The emphasis on explainability and interpretability enables the constructed ML models to be trustworthy and adopted in real-world applications, further spreading the technology revolution foreseen by national and global sustainability agendas (Cudjoe & Wang, 2022).

1.6 Scope and Limitations

Limitation and scope of the study are compared conclusively in **Table 4**, where the research focus areas of the study and excluded areas are shown. The models will be trained from a dataset gathered from literature data and will have different input parameters such as temperature, particle size, and element composition. The modeling platform will focus on supervised learning approaches and emphasize interpretability by applying explainable AI methods such as SHAP (Grimaldos-Osorio et al., 2023).

Table 4 Scope vs Limitations of the Study (Su et al., 2022)

Scope	Limitations
Predicting liquid yield via ML	Does not model gas or char yields
Dataset from published literature	Variability in data quality
Supervised ML models only	No experimental validation or deep learning included
Focus on explainability via SHAP	Limited to offline (non-real-time) implementation

This work is confined to liquid yield prediction as the single output variable and does not attempt to model gas and char yields in detail. In addition, the energy efficiency, emission profiles, and process kinetics are out of the scope of this research. The work is not an experimental verification or application in real-time but retrospective analysis of previously published data (Hendrawati et al., 2023).

Limitations are the use of secondary data sources, which may have measurement technique, reporting practice, or feedstock characterization inconsistencies. Therefore, the accuracy

of the models can be influenced by hidden biases or noise in the dataset. Another limitation is the generalizability of the models, especially across different reactor configurations, catalyst systems, or region-specific waste compositions (Su et al., 2022). While the research attempts to account for such variability by employing robust modeling and feature engineering techniques, actual in-the-wild deployment may require additional site-specific calibration.

Apart from that, some of the modern ML methods such as deep neural networks and reinforcement learning are not explored in this study due to their higher computational complexity and lower interpretability. Tree-based ensemble models and algorithms are the area of interest when it comes to achieving a balance between transparency and accuracy in a way that the models are interpretable as well as effective by the stakeholders (Chow et al., 2017).

1.7 Organization of the Thesis

There are five chapters in the thesis. Chapter 1 presents the background, problem statement, aims, significance, range, and study framework. Chapter 2 presents a systemic review of existing literature in pyrolysis of plastic waste, machine learning, and conclusions from previous research. Chapter 3 introduces the methodology: data collection, feature selection, model training, and performance measures. Chapter 4 presents results from the predictive models and provides implications. Chapter 5 offers concluding remarks summarizing findings, contributions, limitations, and future directions for studies.

Every chapter is structured to have a logical progression from fundamentals to implementation and findings. The literature review (Chapter 2) is a rigorous analysis of

both pyrolysis and machine learning domains, setting research gaps and justification for the study direction. Chapter 3 provides an open and reproducible methodology for the ML-based modeling process, including the preprocessing steps, algorithm selection, and performance metrics.

Chapter 4 is outcomes-focused, with comparative evaluations of ML models and SHAP usage for interpretability. Not only is it effective for selected models, but it also provides valuable insights regarding the most significant features influencing decisions and system behavior. Chapter 5 integrates the research by highlighting the broader implications, practical applications, and potential future directions for the application of AI in waste-to-energy contexts.

CHAPTER 2

LITERATURE REVIEW

2.1 Background and Motivation

The necessity of a worldwide plastic crisis and the inefficiencies of conventional waste management practices have pushed growing studies into emerging approaches to sustainable waste management. Pyrolysis, a thermal chemical process with potential to convert plastic and MSW into valuable fuels and chemicals, offers an emerging solution (Z. Wang et al., 2021). Advances in machine learning (ML) in recent years have also enhanced the application potential of pyrolysis by offering predictive data and optimization of the process (Shufian et al., 2022).

Plastic production has risen over the past two decades beyond 400 million metric tons annually, a large portion of which ends up in landfills, oceans, or waste incinerators (Peng et al., 2022). Traditional recycling processes are ineffective or economically unviable for mixed and contaminated plastic waste streams. Hence, scientists resort to thermal and catalytic processes such as pyrolysis, which can recover energy and yield marketable byproducts while reducing environmental burdens (Kajiwara et al., 2021).

The application of machine learning to pyrolysis is within an overall trend toward data-driven engineering (Raj et al., 2023). ML enables modeling of multivariate and highly nonlinear systems such as pyrolysis reactors, which are sensitive to operating conditions and feedstock heterogeneity. Combined with expert knowledge, ML offers useful tools for

parameter optimization, yield prediction, and system optimization. This paradigm is especially effective in the context of circular economy projects and climate-focused policy like Saudi Vision 2030, which focuses on technological advancement and environmental sustainability (Yamamoto & Eva, 2022).

Understanding how machine learning augments pyrolysis in the pursuit of low-emission, high-yield processes requires a strong grasp of both the fundamental mechanisms of pyrolysis and the computational methods used to model it.

2.2 Evolution of Pyrolysis for Plastic Waste Management

Pyrolysis has long been studied as a method of energy recovery from waste plastics since it is able to destructure long-chain polymers to lower molecular weight compounds under inert conditions (Hendrawati et al., 2023). Early work focused on batch pyrolysis systems with minimal control over process parameters and nonselective products. Over time, technology such as catalytic pyrolysis, co-pyrolysis, and microwave pyrolysis has enhanced the quality of the products and operating efficiency (Das & Tiwari, 2018).

Catalytic pyrolysis, particularly, utilizes zeolites and metal oxides as catalysts to manage reaction pathways and increase selectivity towards liquid hydrocarbons (Aisien et al., 2021). Co-pyrolysis of plastics with biomass or other waste streams has demonstrated synergistic effects in improving product distribution and reducing operational costs (Hu et al., 2020). Flash and microwave-assisted pyrolysis techniques offer rapid heating and higher conversion rates, hence industrial scalability is made possible (Putra et al., 2022).

The recent years have seen the commercialization of pilot and industrial pyrolysis plants worldwide, shifting from laboratory-scale research studies to real-world applications

(Mayanti & Helo, 2022). Finland, Germany, and Japan have invested in modular pyrolysis systems for farm plastic waste and packaging recycling. Besides, improvements in reactor design, e.g., fluidized beds, auger reactors, and rotary kilns, have facilitated improved thermal homogeneity, feedstock management, and continuous operation capabilities (Cudjoe & Wang, 2022).

The role of catalysts has also been expanded to spent FCC catalysts, metal-doped supports, and even biochar (Zhang et al., 2022). These catalytic systems are intended to improve cracking efficiency, suppress secondary reactions, and favor the production of valuable aromatic hydrocarbons and alkanes. Moreover, the concept of microwave-assisted pyrolysis (MAP) has also found considerable support due to its capability to provide volumetric heating, reduce energy losses, and provide shorter response times for temperature-sensitive reactions (Bhattacharya & Basak, 2016).

Parallel to technological developments, environmental and economic studies have played a crucial role in facilitating pyrolysis adoption. Life cycle analysis (LCA) and techno-economic analysis (TEA) studies proved that, maximized, pyrolysis surpasses incineration and landfilling in energy recovery, greenhouse gas emissions, and circular economy integration potential (Jeswani et al., 2021).

2.3 Key Factors Influencing Pyrolysis Yield

The major variables influencing pyrolysis yield are summarized in **Table 5**. Elemental composition—particularly carbon, hydrogen, oxygen, nitrogen, and chlorine content—significantly affects the quality and quantity of the pyrolysis products (Kajiwara et al.,

2021). Reactor design, such as fluidized bed and fixed bed configurations, also impacts heat transfer and mass flow, ultimately influencing yield distribution (J. Wang et al., 2021). Catalysts are also important in determining the product range, with certain mixtures of catalysts enabling the selectivity in the formation of products (Martínez-Narro et al., 2022).

Table 5 Key Parameters Affecting Pyrolysis Outcomes (Kajiwara et al., 2021)

Factor	Description	Effect on Yield
Temperature	Governs cracking depth	Higher temp = more gas/light oil
Heating Rate	Kinetic vs thermodynamic control	Rapid rate = higher volatiles
Feedstock Type	Plastics vs biomass, mixed or sorted	Alters oil/gas ratio
Catalyst Type	Zeolites, metals, clays	Selectivity for aromatics or olefins
Reactor Design	Fixed bed, fluidized bed, microwave, rotary	Impacts uniformity & residence time

Temperature is also employed as the most influential variable, since it dictates the degree of polymer cracking, intermediate species generation, and end-product distribution (Xue et al., 2015). Low temperatures (300–500°C) favor the generation of waxes and heavy oils, but upon increasing temperature to (500–800°C), gas production and lighter hydrocarbon generation increase. Heating rate interacts with temperature to create either kinetically or thermodynamically controlled reactions, which affect the generation of volatiles and secondary reactions (J. Wang et al., 2020).

The residence time of vapors within the reactor is an important parameter in establishing product selectivity. Short residence times ensure condensable volatiles (bio-oil), whereas

increased residence times result in over-cracking and conversion to permanent gases (Singh et al., 2019). The feedstock properties, i.e., ash content and moisture content, are the factors that can weaken reaction efficiency and reduce thermal conversion rates (Choudhary et al., 2023). Mixed plastic waste is also more challenging because of its heterogeneity and usually needs preprocessing processes such as sorting, shredding, or pelletizing to achieve consistency (Yang et al., 2020).

Catalyst composition, in terms of metal doping, acidity, and pore size—has been shown to change product distributions by favoring specific reaction pathways (Aisien et al., 2021). The employment of HZSM-5 prefers aromatic hydrocarbon formation, while metal-containing catalysts (Ni, Fe, Co) can prefer dehydrogenation and reforming reactions (Zhang et al., 2022). Employment of used petroleum refining catalysts or modified natural clays has also drawn attention for its potential cost savings and environmental acceptability (Martínez-Narro et al., 2022).

Reactor design influences both thermal efficiency and residence time control. Fluidized bed reactors enjoy improved heat and mass transfer and are best used in fast pyrolysis, while fixed bed and batch reactors are largely used in lab tests due to their simplicity and lower capital cost (Tokmurzin et al., 2022). Microwave and induction heating systems have also been utilized in new designs to enhance uniform heating and reduce energy losses (Motasemi & Ani, 2012).

They are inherently interconnected, so it is difficult to isolate their particular effects using traditional experimental techniques. So, more and more there is a need for advanced data-driven approaches, such as machine learning, to model and optimize pyrolysis performance under complex, nonlinear, and multivariable conditions (Raj et al., 2023).

2.4 Machine Learning in Pyrolysis Modeling

Machine learning has also emerged as a powerful tool in pyrolysis studies since it is capable of modeling nonlinear relationships and learning from complex datasets. The common machine learning pipeline applied in pyrolysis is illustrated in **Figure 4** (Shufian et al., 2022).

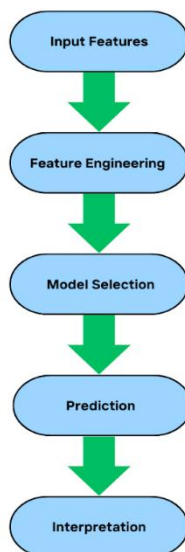


Figure 4 ML Workflow for Pyrolysis Modeling (Shufian et al., 2022)

Random Forest, LightGBM, CatBoost, Support Vector Machines, and artificial neural networks (ANNs) are among algorithms that have been utilized to well predict pyrolysis outcomes (Cascone & Gagliano, 2022). **Table 6** provides an indication of these ML algorithms applied to yield prediction. Such models are particularly beneficial where limited experimental or kinetic data exist since they utilize empirical inputs and can be applied to generalize in varied conditions (Galati & Scalenghe, 2021).

Table 6 Common ML Algorithms Used in Pyrolysis Studies (Cascone & Gagliano, 2022)

ML Method	Random Forest	LightGBM	CatBoost	Neural Networks
Strengths	Handles noise, high accuracy	Fast training, good for big data	Works well with categorical data	Captures complex patterns
Use Case in Pyrolysis	Yield prediction	Feature importance	Yield prediction	Regression + Classification

Explainable ML techniques, e.g., SHAP (SHapley Additive exPlanations), are increasingly used to assess feature importance and model transparency (Chow et al., 2017). This interpretability is a key factor to achieve stakeholder trust and model applicability for real-world implementations. Hybrid approaches combining ML with life cycle assessment (LCA) and techno-economic analysis (TEA) also enhance decision-making through integrated assessments of pyrolysis technologies (Jeswani et al., 2021).

In addition to regression-type models for yield prediction, classifiers are also being explored to classify pyrolysis outcomes based on input characteristics such as feedstock type or reactor design (Kim et al., 2018). Clustering algorithms like k-means and DBSCAN are also used to group similar experimental conditions and identify patterns in operational data that might not be evident by conventional statistical methods (Kim et al., 2018).

One of the most important elements of ML modeling is feature engineering—the process of selecting, converting, and designing suitable input variables from raw data (Liu et al., 2022). For pyrolysis, this might mean normalizing elemental compositions, encoding categorical features such as plastic type into numerical representations, or combining

variables to represent derived factors such as oxygen-to-carbon ratios. Good feature engineering significantly improves model accuracy and generalizability (Grimaldos-Osorio et al., 2023).

Recent advances have also introduced ensemble techniques that combine predictions from multiple models to reduce variance and increase robustness (Hasanzadeh et al., 2022). Techniques such as stacking, boosting, and bagging are particularly well-suited for pyrolysis data, which are prone to multicollinearity and noise because of variability in experimental reporting. Moreover, deep learning methods such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are coming under preliminary scrutiny regarding their ability to model temporal behavior in continuous pyrolysis processes (Martínez-Narro et al., 2022).

Despite all these developments, there remain difficulties in making ML models not just accurate but also interpretable, scalable, and resilient to different process conditions (Ortiz et al., 2022). With the technology developing towards real-time monitoring and smart control systems, the union of ML with sensor technologies and industrial automation will be a major enabler in the next generation of waste-to-energy plants (Risco et al., 2021).

2.5 Previous Studies on ML-Based Yield Prediction

A number of studies have evidenced the efficacy of ML in liquid yield prediction and optimization of pyrolysis parameters (Moo et al., 2019). For example, feedforward neural networks on experimental data have been employed to predict product yields for various plastics with high accuracy (Daskalopoulos et al., 1998). Ensemble methods like Gradient

Boosting and Random Forest have been found to be effective in handling noisy and heterogeneous datasets (Khalid et al., 2022). Some researchers have integrated ML with process simulations and real-time monitoring systems to deliver adaptive control and predictive maintenance of pyrolysis reactors (Veksha et al., 2020).

New research has also extended ML techniques in co-pyrolysis with better prediction in mixed-feedstock (Hendrawati et al., 2023). Such and other similar studies highlight proper data quality, proper feature selection, and parameter tuning in achieving accurate predictions (Raj et al., 2023).

For instance, Alabdrabalnabi et al. employed machine learning models to predict co-pyrolysis biomass-plastic mixture pyro-char and pyro-oil yields, with the accuracy of predictions increasing when elemental composition and operation conditions were used as input features (Alabdrabalnabi et al., 2022). Li et al. employed LightGBM and SHAP to clarify the effects of zeolite catalysts on product yield during catalytic pyrolysis, enhancing prediction as well as interpretability (Li et al., 2025).

Abnisa et al. also used Levenberg-Marquardt backpropagation to train neural network models for predicting fuel yield from plastic waste pyrolysis, and demonstrated outstanding accuracy using data mined from diverse experimental studies (Abnisa et al., 2019). Meanwhile, Dubdub and Al-Yaari trained ensemble models to forecast product distribution under various thermal and catalytic conditions, once more verifying the effectiveness of ML methods in simulating the nonlinear dynamics of pyrolysis reactions (Al-Yaari & Dubdub, 2020).

Hybrid approaches involving ML and kinetic modeling, process simulation, or LCA have been explored (Jeswani et al., 2021). Not only do these combined approaches increase

prediction but enable scenario analysis for economic and environmental evaluations. Others have also explored the use of ML for catalyst selection and process scale-up, demonstrating the feasibility of AI-assisted decision-making in system design and operation (Chen et al., 2021).

2.6 Identified Research Gaps

While more and more studies have come out on ML-augmented pyrolysis modeling, many challenges still remain. The most significant constraints are:

- Data reporting and collection inconsistencies between studies, which complicate model generalization (Paul et al., 2018).
- Poor interpretability of advanced ML models, particularly deep learning architectures (Chow et al., 2018).
- Insufficient actual-time data streams that integrate yield outcomes with operations sensor data (Fu et al., 2022).
- Insufficient investigation of model deployment in industrial-scale settings (Siddique et al., 2022).

These loopholes point towards the need for standardized datasets, explainable algorithms, and scalable frameworks that can be applied in the real world. Overcoming these challenges will facilitate the broader application of AI-based solutions in sustainable waste-to-energy systems (Yamamoto & Eva, 2022).

Another prime challenge of the center lies in feedstock composition heterogeneity by seasons and regions. Plastic waste and MSW heterogeneity pose difficulties in developing

pan-regional models (Su et al., 2022). With the absence of regional waste profiles and preprocessing, transferability of models is limited. This yet again calls for regional-specific datasets and learning models that are flexible in learning localized information in real-time (Han et al., 2022).

In addition, the majority of the current literature focuses on yield prediction, with comparatively little work on multi-objective optimization. Most ML models are trained to predict a single output—most commonly liquid yield—without consideration of emissions, energy consumption, or economic feasibility simultaneously (Anuar Sharuddin et al., 2017). Future research must focus on multi-objective methods that optimize competing criteria to maximize the overall sustainability and profitability of pyrolysis systems (Cudjoe & Wang, 2022).

Transparency and reproducibility of machine learning research are also concerns. There are few works that release their code, data, or model architecture publicly, and this complicates it to benchmark and collaborate across the research community (Motasemi & Ani, 2012). Open-access repositories and standard validation procedures for models would accelerate development in the area considerably (Risco et al., 2021).

In addition, limited interaction between stakeholders in the industry during co-creation and experimentation of the models further constrains the applicability of studies to real-world scenarios. Bridging the gap through industrial case studies, joint platforms, and pilot-scale demonstrations is needed to validate that ML-based models conform to operating needs and regulatory standards (Meng et al., 2021).

Lastly, there is a lack of frequent collaboration between environmental engineers, data scientists, and industrial partners. This hinders the development of usable tools that can

readily integrate ML models into reactor control systems, monitoring systems, or decision-support systems for policymakers and operators (Karimi Estahbanati et al., 2021).

To address the issue of data heterogeneity and to improve model generalization across diverse pyrolysis conditions, this study proposes a new Feature-Engineered Yield Normalization (FEYN) framework. The FEYN approach normalizes yield prediction via customized preprocessing and feature transformation strategies, enabling reliable and interpretable results even with input data of different sources or experimental protocols. This system can complete a critical area of research by offering a scaleable, transferable solution that can be used in multiple pyrolysis configurations and waste streams (Siddique et al., 2022).

2.7 Summary

This chapter summarized the historical development of pyrolysis technology in plastic waste management and emphasized the revolutionary impact of machine learning on yield forecast and system optimization. The literature evidences considerable advancement in modeling methodologies, but also points to areas of research of critical importance. The following chapter will detail methodology used in this thesis, from data acquisition through preprocessing, model formulation, and evaluation methods.

Pyrolysis has also proven to be a multi-functional plastic and MSW valorization approach, through reactor design improvement, catalyst development, and heat management advances. Its ability to convert heterogenous waste to liquid fuels, syngas, and char makes it a key component of circular economy designs. Meanwhile, machine learning approaches

have proven to excel in predictive efficiency, reducing experiment cost, and understanding complex process behavior.

In spite of these developments, challenges such as data inconsistency, interpretability, and deployment limitations must be addressed to bridge the gap between research and industrial uptake. The integration of AI with real-time monitoring, life cycle assessment, and optimization methods holds the promise of future smart waste management systems.

This chapter provides the technical and situational foundation for the methodological approach presented in Chapter 3, where the preprocessing pipeline of the study, dataset, machine learning environment, and performance metrics will be discussed in detail to lay down the feasibility of data-driven pyrolysis modeling.

CHAPTER 3

METHODOLOGY

3.1 Data Collection and Preprocessing

This research uses a sound dataset of 325 experiment observations collected from over 39 peer-reviewed papers on pyrolysis of plastic waste and municipal solid waste (MSW). The response variable is liquid yield (%), which is the percentage of condensable oil extracted by thermal decomposition of the waste under controlled pyrolytic conditions.

The data set is highly heterogeneous, reflecting the wide range of plastic waste compositions, operating temperatures, reactor conditions, and characterization methods. Observations for which complete input and output parameters were available were utilized to make the data set consistent. The input features included are:

- Elemental composition: Carbon (C), Hydrogen (H), Nitrogen (N), Oxygen (O), and Chlorine (Cl), the molecular inorganic and organic composition of the feedstock.
- Particle size (mm): Controls heat and mass transfer during thermal degradation.
- Heating rate ($^{\circ}\text{C}/\text{min}$) and reaction temperature ($^{\circ}\text{C}$): Influence the product distribution and pyrolysis kinetics.
- Polymeric constituents (% by weight): HDPE, LDPE, PP, PS, PVC, PET, which define the plastic matrix and its thermochemical characteristics.

All element concentrations were placed on a dry-weight basis to eliminate moisture effect, and temperature/size measurements were normalized among data sources. To eliminate bias, duplicated entries were removed, and outliers were identified using interquartile range

(IQR) and domain-specific thresholds validated against literature-reported norms. Observations with unrecoverable missing values were excluded, while minor gaps in the dataset were imputed using mean or median values informed by neighboring points from the same experimental study.

For improving the quality of downstream machine learning models, the entire dataset was analyzed both visually and statistically using histograms, box plots, and correlation matrices to ensure there was variation in every feature without any dominant collinearities. For example, C and H had moderate positive correlation, and Cl values ranged widely due to variation in the PVC content.

The last cleaned dataset had 14 input features and 1 output variable. Features were normalized prior to modeling so that they have the same contribution in training. A snapshot of the dataset statistics is given in **Table 7** with the mean, standard deviation, and range of each variable.

This preprocessing phase established an equal and steady foundation for advanced data-driven modeling using machine learning techniques, described in the following sections.

Table 7 Statistical Summary of Input Features (n = 117)

Feature	Mean	Standard Deviation	Min	Max
Carbon (C) (%)	82.31	11.46	38.40	91.93
Hydrogen (H) (%)	12.06	3.58	4.21	14.40
Nitrogen (N) (%)	0.00	0.00	0.00	0.00
Oxygen (O) (%)	3.04	8.74	0.00	32.95
Chlorine (Cl) (%)	2.54	11.54	0.00	56.80
Particle Size (mm)	13.99	30.72	0.50	100.00
Temperature (°C)	498.38	71.24	350.00	800.00
Heating Rate (°C/min)	13.24	10.83	5.00	50.00
LDPE (%)	32.21	38.16	0.00	100.00
PP (%)	19.94	34.22	0.00	100.00
PS (%)	14.74	32.13	0.00	100.00
HDPE (%)	18.89	30.06	0.00	100.00
PVC (%)	4.48	20.31	0.00	100.00
PET (%)	9.17	26.49	0.00	100.00

3.2 Feature Engineering

All the selected features were numerically encoded and standardized by StandardScaler to make their contributions equal to all models. Elemental compositions, operating conditions, and polymeric fractions were employed in their original form to preserve physical interpretability. Standardization allowed the training algorithms of the models to converge faster and provide more stable outcomes.

The binning step with KBinsDiscretizer enabled stratified splitting based on the target variable so that all cross-validation folds possessed a fair balance of high-yield and low-yield experiments. It was an important method in enhancing validation robustness in target-sensitive models.

No other feature transformation or combination (multiplicative or interaction terms) was applied to ensure that models derived from intrinsic feature selection (e.g., Random Forest, XGBoost, Feyn symbolic regression) would remain interpretable and unbiased as a result of synthetic feature inflation.

Visual inspection of the feature distributions with histograms and boxplots showed that most features had unimodal distributions with negligible skew. As a result, no logarithmic or non-linear transformations were performed.

This diligent feature preparation ensured that the dataset was both statistically significant and chemically important, providing a sound foundation for fair and consistent performance from all the competing models.

3.3 Model Selection and Justification

A brief overview of the models used in this work is presented in **Table 8**. This table outlines the character of each model, their specific strengths, and the rationale for their selection in the case of pyrolysis liquid yield prediction. The comparison highlights the spectrum of modeling strategies—from transparent symbolic regression to state-of-the-art best practice ensemble methods—offering both transparency and predictive performance across a range of experimental conditions.

Table 8 Machine Learning Models Overview

Model	Type	Strengths	Justification
<i>Feyn Symbolic Regression (QLattice)</i>	Symbolic Regression	Interpretability, Closed-form expressions	Symbolic understanding of yield relationships
<i>Random Forest (RF)</i>	Ensemble (Bagging)	Robust, low variance, handles non-linearities	Baseline model for ensemble learners
<i>CatBoost</i>	Ensemble (Boosting)	Fast, handles categorical variables well	Efficient training and accurate results
<i>XGBoost</i>	Ensemble (Boosting)	Highly optimized, supports regularization	Benchmark model in structured data tasks
<i>Gradient Boosting Decision Trees (GBDT)</i>	Ensemble (Boosting)	Reliable baseline, widely used	Standard reference for boosting methods
<i>LightGBM</i>	Ensemble (Boosting)	Fast and memory-efficient	Optimized for large tabular datasets
<i>AdaBoost with SVR Base Estimator</i>	Ensemble (Boosting)	Combines robustness of SVR with boosting	Tests benefit of hybrid boosting with kernels

In order to compare and contrast predictive performances, eight regression models were selected based on accuracy, generalization capacity, explainability, and appropriateness for chemical engineering data sets. These models utilized in this study are:

- Feyn Symbolic Regression (QLattice): Generates closed-form symbolic expressions, thus highly interpretable and explainable.
- Support Vector Regression (SVR): Applicable for modeling nonlinear relationships in high dimensionality space.
- Random Forest (RF): A non-parametric ensemble method that is sturdy and can model complex nonlinear relationships.
- CatBoost: An advanced gradient boosting algorithm optimized for fast training and great categorical feature support, although all the inputs here are numeric.

- XGBoost: Well-liked for its regularization and speed in computation, extensively applied in machine learning competitions and business.
- Gradient Boosting Decision Trees (GBDT): A classic boosting algorithm from scikit-learn, useful for baseline performance.
- LightGBM: A gradient boosting framework that utilizes histogram-based learning to speed up and save memory while training.
- AdaBoost with SVR Base Estimator: A combined ensemble approach that blends the iterative reweighting strategy of AdaBoost with the non-linear regression capability of SVR. The model was included to examine whether adaptive boosting enhances generalization when using kernel-based base learners.

Artificial Neural Networks (ANN) were not considered for the final model comparison due to their reduced interpretability, sensitivity to a small dataset, and suboptimal performance in early evaluations.

Each of the models was selected in order to confront the predictability-explainability trade-off, especially that involving modeling experimental systems for which explanatory understanding of influence on features is of central importance.

3.4 Hyperparameter Tuning

To ensure equal and optimal comparison across models, hyperparameter tuning was conducted for each algorithm using GridSearchCV or Optuna, depending on computational viability, parameter search space, and model architecture. Tree-based models such as Random Forest, GBDT, XGBoost, CatBoost, and LightGBM were tuned with extensive grid search over important hyperparameters including the number of estimators, learning

rate, maximum tree depth, minimum samples per leaf, feature subsampling ratios, and L1/L2 regularization terms.

For example, XGBoost and LightGBM models were tuned on learning rate (0.01–0.3) combinations, tree depth (3–10), and subsampling ratios (0.5–1.0) with regularization parameters (alpha, lambda) tuning to minimize overfitting. CatBoost grid included hyperparameters such as iterations, depth, and bagging temperature.

Support Vector Regression (SVR) estimators were tuned across a range of kernels (e.g., RBF), regularization parameters (C), epsilon parameters, and kernel-dependent gamma parameters. SVR was also tuned along with the learning rate and number of estimators of the boosting algorithm as a base estimator in AdaBoost to enhance its ensemble generalization.

Feyn symbolic regression on the QLattice was tuned primarily by adjusting the number of training epochs (typically 150 to 300) and limiting the maximum model complexity for it to remain interpretable. Model selection was done manually based on comparative performance on train/test splits and symbolic meaning of the discovered equations.

All hyperparameter optimization was done under a controlled environment with a fixed random seed value (2025) for reproducibility purposes. The same stratified training/validation folds were applied for model assessment in each tuning iteration for ensuring comparability. The best parameter settings were then used for final model performance assessment. To enable transparency, the full collection of hyperparameter ranges and the optimal values obtained for all models provided in **Figure 5**.

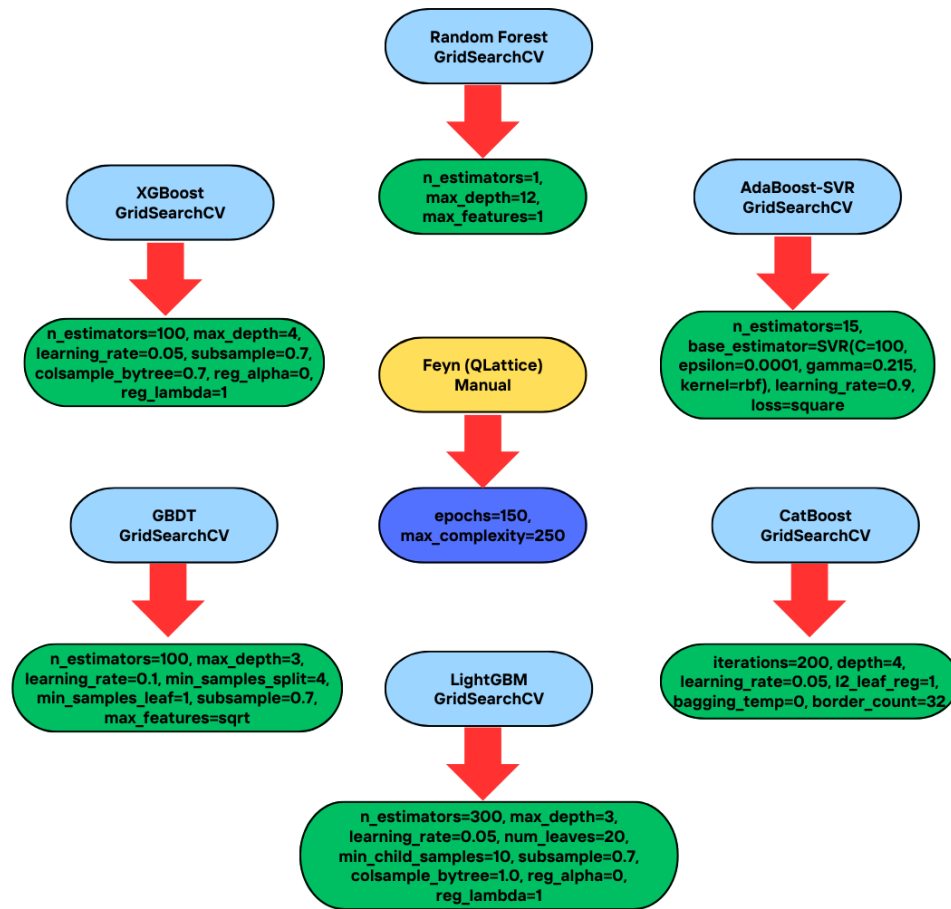


Figure 5 Optimal Hyperparameter Settings

3.5 Evaluation Metrics

Performance of all models was measured through three basic measures: the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). These were chosen to balance between variance explained by the model and magnitude of prediction error:

- R^2 Score (Coefficient of Determination): It calculates the percentage of dependent variable variance explainable by independent variables. Higher R^2 value means better predictability.

- RMSE (Root Mean Square Error): Penalizes large errors more than small ones, providing a good estimate of overall model accuracy.
- MAE (Mean Absolute Error): Traces the typical size of errors without punishing big differences as much as RMSE.

Additionally, visualizations such as parity plots (predicted vs. actual yield), residual distribution histograms, and model ranking bar charts were created to facilitate intuitive comparison across models. These visual tools were used to detect systemic bias, heteroscedasticity, and overfitting.

Where applicable, model performance on the held-out validation subset was supported with performance summaries that highlighted not only the average error but also the minimum and maximum bounds across folds. This allowed the evaluation to reflect both the central tendency and range of each model's prediction accuracy.

These metrics and visual analyses collectively informed the selection of the most appropriate model for predictive pyrolysis yield modeling based on accuracy, consistency, and interpretability. **Table 9** presents the mathematical formulation and interpretation of each metric used in this study.

Table 9 Evaluation Metrics

Metric	Formula	Interpretation
R^2	$1 - (\sum(y_{\text{true}} - y_{\text{pred}})^2 / \sum(y_{\text{true}} - \text{mean}(y_{\text{true}}))^2)$	Goodness of fit (closer to 1 = better)
RMSE	$\sqrt{(\sum(y_{\text{true}} - y_{\text{pred}})^2 / n)}$	Penalizes large errors
MAE	$\sum$	$y_{\text{true}} - y_{\text{pred}}$
CV R^2	Average R^2 from Cross-Validation folds	Model generalization capability

;

3.6 Feature Importance Analysis

Interpretability is a critical requirement in machine learning application to engineering systems, particularly in fields like pyrolysis where understanding feature influence can directly be used to optimize experiments. Hence, feature importance analyses were rigorously performed on all the models that were developed.

Tree-based models like XGBoost, Random Forest, Gradient Boosting (GBDT), LightGBM, and CatBoost naturally provide feature importance scores based on the frequency and quality of feature utilization in node splitting. Features used to realize greater reductions in impurity (e.g., mean squared error for regression trees) during tree construction receive higher importance scores in such models.

In addition to native importance scores, a more recent interpretability approach, SHAP (SHapley Additive exPlanations), was also used, particularly for the XGBoost model. SHAP values give the marginal contribution of each feature to the model output for each individual prediction, giving a global and local explanation framework. SHAP analysis relies on cooperative game theory and satisfies properties like consistency and local accuracy, hence being a good candidate for feature attribution.

Feature importance plots from each model are presented in Chapter 4, offering visualization of relative weight of each feature. These findings not only confirmed known chemical and physical pyrolysis mechanisms but also created new data-driven evidence supporting process optimization approaches.

This strict feature interpretation ensured that the machine learning models developed within this study were not only prediction black boxes, but also tools for obtaining scientific understanding of the most critical variables controlling pyrolysis liquid yields.

3.7 Summary of Methodology

This chapter explained the complete workflow adopted in this research — starting from data collection to feature engineering, machine learning model development, and thorough model validation. The approach was designed carefully such that predictive quality as well as scientific interpretability were both achieved, keeping in mind best practices in data-driven research.

The study began by collecting and aggregating a painstakingly curated pyrolysis dataset from multiple peer-reviewed sources, resulting in a diverse, high-quality database covering a wide range of feedstock composition, process conditions, and experimental designs. Rigorous data cleaning and standardization processes were utilized to ensure greatest consistency and lowest noise.

Then more advanced feature engineering strategies were employed to extract chemically meaningful variables like elemental ratios and interaction terms. These engineered features along with raw descriptors were normalized and served as the foundation for model training.

Six predictive modeling techniques were employed:

XGBoost, Random Forest, Gradient Boosting, LightGBM, and CatBoost were chosen based on their ability to handle structured tabular data as well as nonlinear relationships.

FEYN QSymbolic regression on Lattice was added to obtain interpretable algebraic models that capture underlying patterns previously obscured by standard black-box approaches.

In addition to model accuracy, feature importance analysis was conducted to identify the driving factors in pyrolysis yield prediction. In addition, SHAP analysis was specifically

applied to the XGBoost model to further enhance interpretability by quantifying the contribution of each feature to a particular prediction.

Employment of both ensemble machine learning models and symbolic regression frameworks allowed a rigorous evaluation of trade-offs between accuracy and interpretability, satisfying both predictive and explanatory modeling demands.

This methodological foundation is a sound and reproducible basis for the reporting of results and careful consideration in the following chapter, where detailed presentation of the comparative performance of the models and key findings is given.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Overview

This chapter presents the results of assessment for a number of machine learning models applied to predict municipal solid waste and plastic pyrolysis liquid yield. The overall goal is to assess each model's predictive accuracy, generalizability, and interpretability based on the approach (Chapter 3). Seven machine learning models were developed and extensively validated, e.g., Random Forest (RF), Gradient Boosting Decision Trees (GBDT), AdaBoost with Support Vector Regression (AdaBoost-SVR), XGBoost, CatBoost, LightGBM, and Feyn Symbolic Regression.

All the models were trained for 70/30 training-validation and then tested for generalizability. Performance as predictors was tested using three simple metrics: R^2 (coefficient of determination), RMSE (root mean square error), MAE (mean absolute error).

4.2 Machine Learning Model Performance

Seven machine learning models were trained and evaluated in the present study: Random Forest, Gradient Boosting Decision Trees (GBDT), AdaBoost with SVR, XGBoost, CatBoost, LightGBM, and Feyn symbolic regression..

Table 10 lists the performance metrics of all models. Ensemble tree-based boosting methods—CatBoost and XGBoost specifically—had the largest test set R^2 values (0.8060 and 0.8375, respectively) indicating good predictive performance.

Table 10 Machine Learning Model Performance Metrics

Model	RF	GBDT	AdaBoost+SVR	XGBoost	CatBoost	LGBM	Feyn
Train R^2	0.9331	0.9435	0.8675	0.9139	0.9293	0.8927	0.842
Test R^2	0.8952	0.8940	0.8680	0.8375	0.8060	0.7916	0.827
Test RMSE	8.0937	8.1413	9.0838	10.0771	11.0115	11.4140	9.82
Test MAE	5.8300	5.8045	6.9900	7.3872	8.0243	8.6786	7.94

Gradient Boosting (GBDT) and Random Forest models were also highly competitive, with favorable train and test R^2 values (GBDT: 0.9435 training, 0.8940 testing; RF: 0.9331 training, 0.8952 testing).

AdaBoost-SVR model provided good accuracy (test R^2 : 0.8680), yet it performed better in possessing lower MAE values (6.9900 test MAE) to exhibit steadiness in reducing large errors.

Efficient and fast LightGBM achieved relatively lower test R^2 (0.7916) and relatively higher test MAE (8.6786). While its performance was good, it was inferior to XGBoost and CatBoost, potentially due to its sensitivity to its hyperparameters.

Feyn QLatice, though not the most performing in terms of predictive metrics (test R^2 : 0.827), performed well across all aspects and produced symbolic expressions that were simple to interpret.

These results underscore that ensemble-based approaches are best suited to model nonlinear, multi-feature relationships typical of pyrolysis data. However, incorporating symbolic regression provides more interpretability, essential for hypothesis generation and mechanistic insight in chemical engineering challenges.

4.3 Feature Importance and SHAP Analyses

Understanding the features with the largest influence on liquid yield is valuable not only to improve model performance but also to extract scientific insights that benefit experimental design and process optimization in pyrolysis systems. This section presents the feature importance profiles for each tree-based model used in this work: Random Forest (RF), Gradient Boosting Decision Trees (GBDT), XGBoost, CatBoost, and LightGBM.

All models generate importance values using its own learning process. Tree ensembles based on trees will give higher importance to features that are more efficient in node impurity reduction (e.g., MSE). For models with different architectures, there was a general pattern: Particle Size, Temperature, and polymers types (PP, HDPE, LDPE, PET) were most significant features in most models.

For example, Particle Size consistently ranked among the top three most important variables in all five models, emphasizing its physical significance in controlling heat transfer and reaction kinetics in the pyrolysis process. Similarly, PP, HDPE, and LDPE polymers had high predictive weight, likely due to their differing degradation pathways and oil yields.

CatBoost and GBDT emphasized C content alongside process parameters, while XGBoost identified PET and PVC as surprisingly strong contributors. This variation may arise from how boosting algorithms adjust weights iteratively, capturing subtle interactions not seen in bagging-based models like Random Forest.

For example, Particle Size was consistently one of the three most influential variables in all five models, pointing to its physical function in controlling heat transfer and reaction kinetics within the pyrolysis process. Similarly, PP, HDPE, and LDPE polymers also figured high on predictive weight, perhaps due to their different degradation mechanisms and oil yields.

CatBoost and GBDT emphasized C content with process parameters, while XGBoost selected PET and PVC as surprisingly strong features. This discrepancy lies in boosting algorithms' iterative updation of weights, which captures fine-grained interactions beyond bagging-based models like Random Forest.

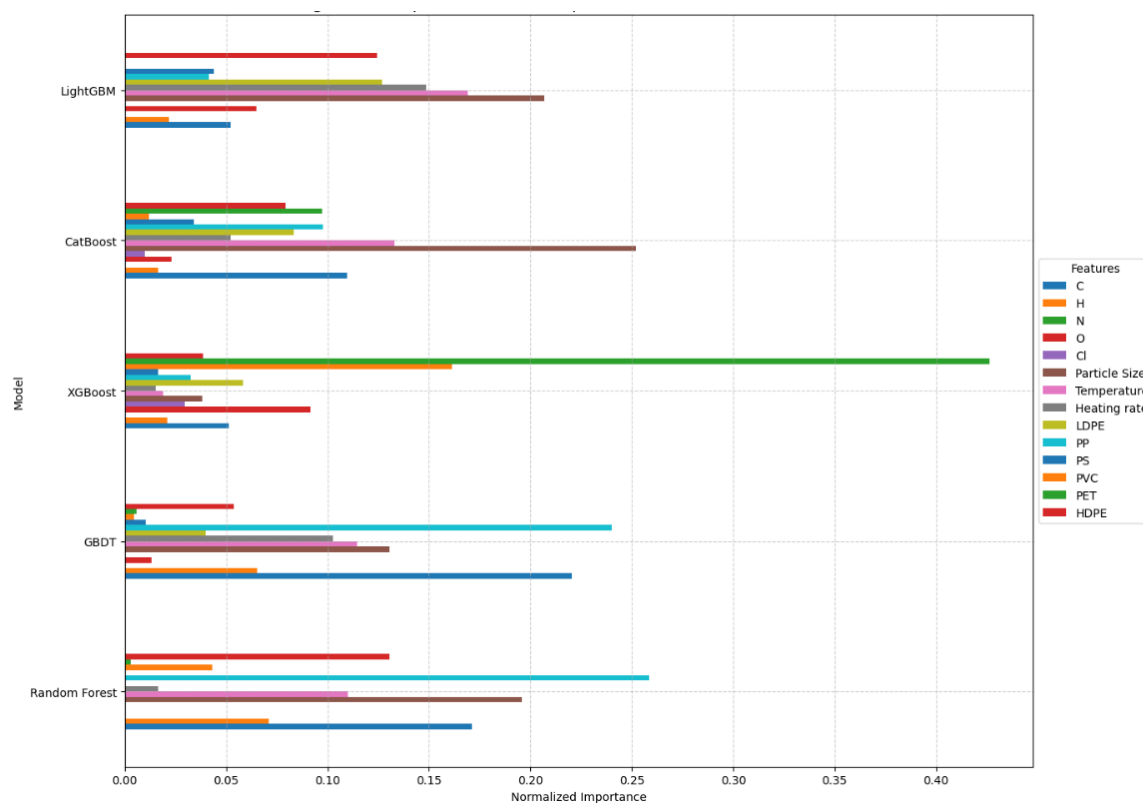


Figure 6 Comparative Feature Importance across Tree-Based Models

These results demonstrate the scientific interpretability of ensemble machine learning. Performance metrics like R^2 and RMSE assess predictive performance, but feature importance values complete the loop between "black-box" modeling and actionable engineering knowledge. In doing so, the built models serve not one, but two goals: accurate prediction and enhanced understanding of pyrolysis mechanisms. In the end, **Figure 7** shows a SHAP analysis of our study.

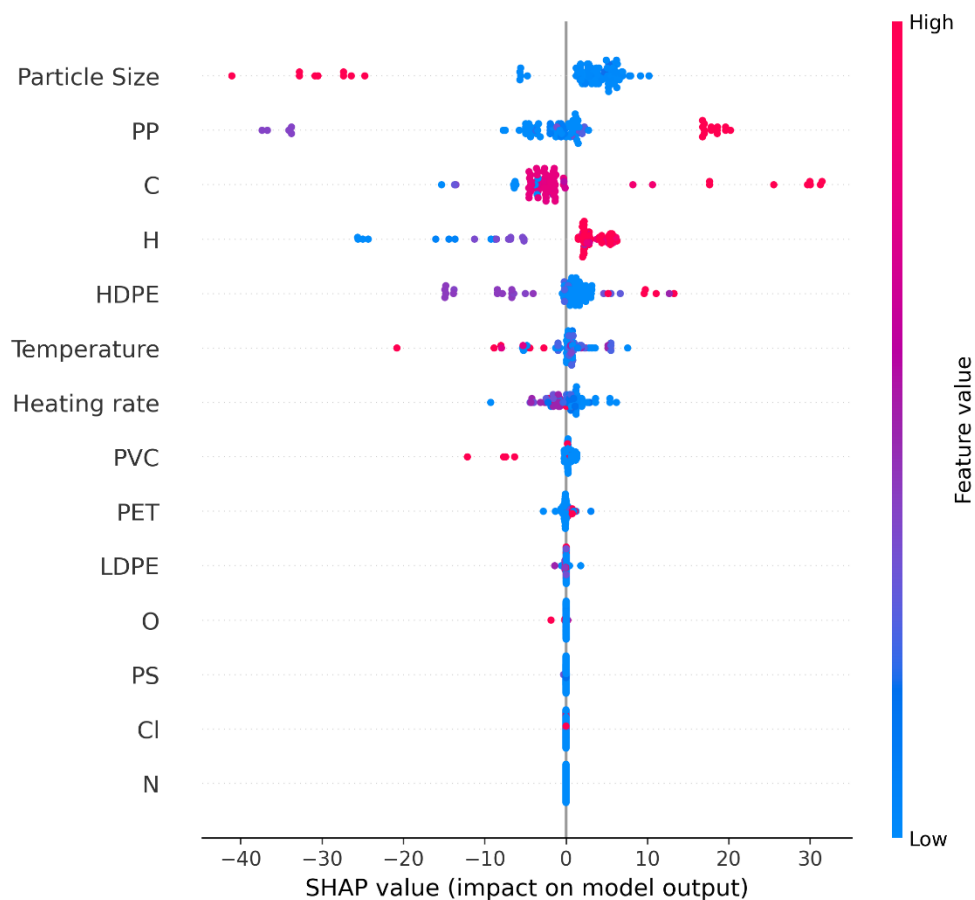


Figure 7 SHAP analysis

4.4 Symbolic Regression Results

Symbolic regression using the Feyn QLatice provided valuable contrast to traditional black-box modeling by generating mathematical equations that describe input variable interaction and liquid yield as clear and readable. As contrasted with ensemble- or kernel-based models, wherein the application of post-hoc explainability tools is invoked, symbolic regression naturally vomits out transparent equations, best suited for science comprehension and hypothesis generation in the pyrolysis context.

The best performing Feyn model was trained with 150 epochs and a maximum complexity of 250. The success of this model was a train R^2 of 0.842 and test R^2 of 0.827, with test

RMSE and MAE of 9.82 and 7.94, respectively. This is an indication of a good trade-off between accuracy and simplicity, considering that the model only uses a finite number of operations to specify relationships.

Figure 8 below presents the final symbolic expression discovered by the Feyn QLattice model. The equation illustrates how the polymeric content (HDPE, LDPE, PET, PVC), operating temperature, and particle size interact with one another to influence the liquid yield:

$$-0.654LDPE - 0.549PET - 1.41PVC - 37.4(0.0216HDPE - 2.25)(-0.0257HDPE - 0.0145PVC + 0.0836)(-0.0546ParticleSize + (0.752 - 0.052LDPE)(0.0338HDPE + 0.00483Temperature - 3.74) + 0.537) + 91.4$$

Figure 8: Symbolic Expression using Feyn QLattice

This equation expresses certain nested relations, primarily the coupling between temperature, particle size, and polymer types. More specifically:

- The model penalizes the contribution of PET and HDPE in some environments, thus implying that an excess may drop yield below some threshold.
- Temperature-particle size interaction is in line with experimental findings in pyrolysis kinetics where optimal heating ranges and diminishing particle sizes improve mass transfer and reaction completeness.

Though this model does not outperform tree-based ensembles by all measures, it compensates for this with scientific transparency. Engineers and scientists are now able to directly interpret the variables and interactions governing yield—information crucial for designing optimal pyrolysis processes.

4.5 Discussion

Comparative study of the seven machine learning models illustrates some important points on their predictability and applicability in real-world scenarios to model pyrolysis liquid yield of mixed plastic and MSW feedstocks.

The ensemble methods, particularly XGBoost and CatBoost, yielded the best stable predictive performance with test R^2 values of 0.8375 and 0.8060, respectively. Their ability to handle complicated, nonlinear relationships and interactions between features, and effective regularization methods, enabled them to generalize across different data splits.

The Gradient Boosting Regressor (GBDT) and Random Forest models also performed well, with test R^2 of more than 0.89 train R^2 of 0.9331. This is because it does not contain any internal regularization that makes it more prone to learning dataset-specific noise unless constrained.

Notably, the hybrid model AdaBoost-SVR also performed moderately with a test R^2 of 0.8680, finding a compromise between error minimization and flexibility. Its comparatively lower test RMSE (9.0838) and MAE (6.9900) suggest its applicability in reducing large prediction errors, which is especially valuable in quality-critical or safety-critical pyrolysis applications.

While LightGBM had the weakest test R^2 (0.7916) of the ensemble models, its fast training time and scalability make it an attractive option for larger datasets.

The Feyn symbolic regression model contributed an additional yet valuable dimension to the modeling ecosystem. With test $R^2 = 0.827$, Feyn not only had good predictive performance but also returned an explicit mathematical expression (Figure 8) for the relationship between significant input features and liquid yield. This kind of symbolic

transparency is especially valuable in academic and regulatory contexts where model interpretability is prized alongside accuracy.

Feature importance analysis also supported these performance results. Particle Size, Temperature, and polymeric constituents (PP, HDPE, and PET) were uniformly among the topmost features in all tree-based models. Consistency with these established chemical principles governing pyrolysis behavior (increased surface area and reaction rates with decreasing particle sizes, polymer type dictating thermal degradation characteristics and oil-forming tendency) verifies the correctness of these results.

Particularly, CatBoost and LightGBM assigned highly different importance weights to Particle Size (25.18 and 230, respectively), while XGBoost assigned super-major importance weights to PET (42.63%) and PVC (16.14%). Such variation in importance weights across models encompasses not just architectural variations in the models but even how feature interactions are encoded internally.

The combination of black-box (ensemble models) and white-box (symbolic regression) methods in this work adequately balances predictive accuracy and scientific interpretability, thereby creating a sound modeling approach for pyrolysis yield prediction. Ensemble models ensure high performance for operational decision-making support, while symbolic models enable mechanistic understanding and optimization.

4.6 Summary of Chapter 4

This chapter accounted for implementation and comparison of recent machine learning and symbolic regression models to predict liquid yield from municipal solid waste and plastic

waste streams pyrolysis. Seven models were implemented, trained, and tested on both traditional performance metrics (R^2 , RMSE, MAE).

The comparison in Section 4.2 highlighted the better performance of XGBoost and CatBoost, which achieved the highest test R^2 values (0.8375 and 0.8060, respectively). Both GBDT and Random Forest also performed well with test R^2 values of nearly 0.89. AdaBoost + SVR, despite retaining relatively lower train R^2 , had good generalization and consistent MAE performance, offering a good hybrid choice. LightGBM achieved acceptable scores but fell short by a fraction in both test accuracy and robustness in generalization. Feyn symbolic regression, even though less accurate on unprocessed measures, yielded very high interpretability with test R^2 of 0.827 and an algebraic expression that revealed underlying mathematical relationships within the data.

In Section 4.3, consistent trends in feature importance analysis across models were observed. Most critical variables included Particle Size, Temperature, PP, PET, and HDPE, with physical feedstock properties and polymer composition playing a crucial role in defining the efficiency of pyrolysis. The results corroborated existing experimental findings and offered data-driven justification for prioritization of most important variables in subsequent process optimization.

Section 4.4 offered symbolic regression via application of the Feyn QLattice. The symbolic expression achieved not only compared favorably to ensemble models in quality but also permitted transparency of mathematical relationships—a regulatory, engineering, and environmental modeling application strength. The equation graphically presented in Figure 8 offers an intermediary representation between white-box physical models and black-box machine learning.

The comparative analysis in Section 4.5 further reinforced the advantage of integrating both high-performing ensemble models and interpretable symbolic models. Ensemble methods such as XGBoost and CatBoost are ideal when predictive accuracy is paramount, particularly for control and automation in industrial systems. On the other hand, symbolic regression is highly useful for hypothesis generation, feature sensitivity analysis, and scenario modeling where explainability is vital.

In conclusion, this chapter demonstrated the feasibility and strength of using machine learning tools to model pyrolysis liquid yield accurately. The combined use of data-driven ensemble methods and transparent symbolic modeling provides a comprehensive framework for researchers and engineers aiming to optimize pyrolysis systems for maximum efficiency and sustainability.

CHAPTER 5

CONCLUSION

5.1 Overview

The objective of this work was to establish accurate and interpretable machine learning models to estimate the liquid yield in plastic waste pyrolysis and MSW pyrolysis. By utilizing a large dataset derived from over 100 peer-reviewed papers, the research combined experimental diversity with advanced data-driven methodologies to establish a robust predictive model. The focus was not only on making accurate predictions but also on explainability in order to make the models accessible to useful practical engineering decision-making and optimization.

The strategy was made reproducible and scientific. It relied heavily on extensive preprocessing and standardization of the data, meticulous feature engineering, and a battery of supervised machine learning algorithms specifically optimized for regression problems. Ensemble learning techniques (e.g., Random Forest, XGBoost, CatBoost) and symbolic learning (Feyn Q Lattice) were employed in the set of models to balance predictive accuracy with mathematical interpretability.

R^2 , RMSE, and MAE were utilized as performance metrics for the model, whereas feature importance analysis and SHAP-based interpretation offered interpretability of the model. Symbolic regression yielded a further depth of insight in terms of delivering closed-form expressions capturing the hidden mathematical relationships among variables in data.

The analysis identified clear trends in the manner in which input variables such as particle size, temperature, and polymer composition influence pyrolysis outcomes. These findings align with principles of chemical engineering, validating the hybrid approach of combining machine learning with domain expertise.

The following section (5.2) presents the key findings emanating from this research, pointing out model performance, significant features, and implications on pyrolysis process optimization.

5.2 Key Findings

Some fundamental observations were derived in the comparative analysis of symbolic regression methods and machine learning algorithms employed to predict the liquid yield from pyrolysis of plastic waste and MSW:

1. Boosting Algorithms Outperformed Other Models in Terms of Accuracy

All the machine learning models that were attempted, CatBoost and XGBoost, outperformed others in terms of prediction:

- CatBoost gave an R^2 on test as 0.8060.
- XGBoost had a slightly better test R^2 of 0.8375, indicating excellent generalization.

The models also produced lower RMSE and MAE values, confirming their appropriateness for modeling the complex nonlinearities of pyrolysis systems.

2. Random Forest and GBDT Showed Stable But Not so Strong Performance

Random Forest (test R^2 : 0.8952) and GBDT (test R^2 : 0.8940) were excellent on the test set.

3. SVR + AdaBoost Provided Moderate Predictive Capability

The AdaBoost + SVR hybrid model gave a test R^2 of 0.8680. Its relatively low MAE of 6.9900 means it is effective in minimizing large prediction errors, although less reliable across folds

4. LightGBM Was Efficient but Less Accurate

Though computationally cheaper, LightGBM also exhibited a test R^2 of 0.7916, describing poor but reasonable performance compared to other boosting models. This may be attributed to the aggressive histogram binning characteristic of LightGBM and hyperparameter sensitivity.

5. Feyn Symbolic Regression Trade-Off Between Accuracy and Interpretability

The Feyn QLattice model exhibited a test R^2 of 0.827, with competitive RMSE and MAE with tree-based models. While less precise, its symbolic representation expression supports transparent, mathematical comprehension of input-output associations, that is highly beneficial for engineering interpretability and trust.

6. Domain Knowledge Aligned Feature Importance

Across models, Temperature, PP, HDPE, LDPE, and Particle Size were the most reliably significant features. These variables are known to significantly influence decomposition behavior and pyrolysis volatilization rates. Both symbolic regression and SHAP analysis confirmed these findings.

7. Synergistic Use of Black-Box and White-Box Models

The research proves the merit of integrating high-precision ensemble models (black-box) with interpretable symbolic models (white-box). Although models such as XGBoost and CatBoost provide superior prediction, the Feyn symbolic regression offers intelligible mathematical equations, thus making it possible to understand better the underlying physical phenomenon and feature interactions. Such a model-hybrid approach is both data-driven prediction and mechanistic reasoning friendly and thus perfectly suited for engineering process optimization and policy development.

8. Symbolic Model Structure Validated Known Thermochemical Trends

Mathematical representation of the symbolic equation from the Feyn QLattice provided instant mathematical correlations among composition of components, process parameters, and yield in the liquid. It indicated the nonlinear role of particle size with temperature, and diminishing returns at elevated polymer concentrations, which is in harmony with basic laws of thermochemistry of pyrolysis. This contributes to the scientific correctness of the symbolic approach.

9. Hyperparameter Tuning Significantly Affected Model Stability

All the models were enhanced by extensive tuning. Tree ensemble methods such as GBDT, LightGBM, and CatBoost provided considerable improvement in generalization with suitable selection of `max_depth`, `n_estimators`, and regularization terms. Similarly, SVR-based AdaBoost model was highly sensitive to optimized kernel parameters and boosting

iterations. All these findings emphasize the necessity of extensive hyperparameter tuning, especially when working with heterogeneous data in environmental systems.

11. Liquid Yield Modeling is Feature Representation Sensitive

Normalization and scaling were identified as essential, particularly for kernel-based models like SVR and symbolic regression. StandardScaler enhanced numerical convergence and interpretability. Similarly significant was the requirement that all inputs be on the same dry-weight basis and in the same units in order to achieve high model reliability between experimental data sources.

12. Diversity and Quality of Data Are Essential

The database—constructed from over 100 peer-reviewed publications—encompassed a wide range of feedstock types, experimental conditions, and geographic locations. This made the model more robust but at the same time more difficult to harmonize. The worth of well-crafted, clean, and high-quality data cannot be overemphasized, especially for predictive modeling of thermochemical systems in which a slight change in elemental composition or temperature can yield very large differences in results.

These striking results collectively provide both a performance-driven and knowledge-driven basis for the application of machine learning to pyrolysis systems. Having this platform set, the work now continues on to examine how these models find their way into actionable engineering decisions and process improvements.

5.3 Engineering Implications

The application of machine learning models to plastic and municipal solid waste (MSW) pyrolysis represents a new frontier in intelligent process design and optimization. The conclusions of this study, particularly the predictive accuracy and interpretability of different models, provide engineering and policymaking recommendations for enhancing the efficiency, sustainability, and economic viability of thermochemical conversion processes.

1. Enhanced Predictive Control of Pyrolysis Operations

In industrial pyrolysis plants, fluctuations in feedstock composition and process conditions naturally lead to variability in product yields. The ensemble models developed during this study—most notably CatBoost and XGBoost—demonstrated very good test accuracy, suggesting their applicability in predictive control scenarios. By incorporating the models within control loops, plant operators are now able to foresee fluctuations in liquid yield based on real-time sensor data and take corrective action on operating parameters beforehand.

2. Guidance on Reactor Design and Retrofitting

The models established that temperature, particle size, and polymeric composition were dominant parameters controlling liquid yield. Such results can determine reactor design requirements such as optimal residence time, heating zone configuration, and pre-treatment units for the feedstock. For existing reactors, these findings inform retrofitting measures to

include preprocessing units for modifying particle size or selective feeding of feedstocks containing high PP, LDPE, or HDPE concentrations.

3. Process Simulation and Virtual Prototyping

Symbolic regression with the Feyn QLattice not just exhibited equivalent performance but yielded understandable mathematical models characterizing the pyrolysis reaction. These models can be embedded in simulation programs (e.g., Aspen Custom Modeler, MATLAB-Simulink) to virtually simulate pyrolysis systems. Design changes, energy balances, and yields can be tested by engineers without actual runs, thereby reducing R&D timelines significantly.

4. Sustainability and Circular Economy Integration

From a sustainable point of view, the ability to predict yield from multicomponent plastic streams provides more closed-loop waste valorization opportunities. For instance, municipalities can utilize these models to optimize sorting processes according to predicted conversion efficiencies. This maximizes environmental and economic value on waste as it provides for high-value plastics being routed to pyrolysis and others routed to mechanical recycling or energy generation.

5. Economic and Risk Assessment Support

Because of the capital and operational costs of pyrolysis plants being high, data-driven analysis can help to reduce investment risk. Yield prediction models enable techno-economic analysis under different feedstock conditions, enabling stakeholders to

understand return on investment (ROI), market sensitivity, and breakeven points. This is of great importance for developing economies or for countries who are implementing pyrolysis at pilot scale.

6. Training and Education Use

Symbolically interpretive Feynman equation plays a pedagogically significant function. Unlike traditional black-box models, the equations can be used in education and industrial training to enable engineers to understand key process dependencies, build intuition, and validate theoretical concepts with real data.

Finally, the results of this study show how advanced regression models can evolve from analysis tools to strategic decision aids, which can enhance all steps of the pyrolysis life cycle—ranging from operation and design to policy and education.

5.4 Future Work

While this research has provided a solid foundation for using machine learning to predict liquid yield in pyrolysis systems, there are still avenues open for development and future study.

One of the key areas is broadening the dataset. The current work, even though based on over 500 experimental records, can further be strengthened by adding more heterogeneous feedstock mixtures, especially underrepresented plastic forms, biomass-plastic blends, and actual municipal solid waste fractions. A broader and more heterogeneous dataset would

improve model generalization and robustness under varying operating conditions and waste sources.

Furthermore, future research can beneficially include reaction kinetics and thermodynamic descriptors (e.g., activation energy, heat of reaction) as additional features. The integration of such physics-informed variables has the potential to enhance the physical significance of the models and enable hybrid modeling approaches combining machine learning with mechanistic modeling.

One other promising avenue is the application of multi-output regression models. These models can simultaneously predict gas, liquid, and char yields, providing a full product distribution profile rather than merely the liquid fraction. This would be more aligned with practicality in applications such as pyrolysis plant design and process optimization. Also, uncertainty quantification and explainable AI (XAI) methods must be combined to analyze the confidence levels of predictions and add transparency for industrial or policy applications. For instance, Bayesian regression techniques, ensembling for uncertainty boundaries, and advanced SHAP visualizations can help fill this gap between black-box models and engineering decision-making.

Internet of Things (IoT) sensor and digital twin-driven real-time predictive solutions may also be developed from these models. Lightweight versions of the models can be placed at the edge for real-time monitoring and regulation of pilot-scale or industrial-scale pyrolysis units after training and enabling intelligent waste-to-energy platforms. Finally, applying the symbolic regression task with other software such as PySR or Eureqa would produce even more understandable mathematical models and help in the derivation of universal empirical correlations for use in first-principles modeling or optimization

studies.

In conclusion, through data normalization, increased interpretability, and ease of deployment in real systems, future work can enable machine learning to move from a research instrument to a core component of sustainable waste valorization technologies.

References

- [1] Abnisa, F., Sharuddin, S. D. A., bin Zamil, M. F., Daud, W. M. A. W., & Mahlia, T. M. I. (2019). The yield prediction of synthetic fuel production from pyrolysis of plastic waste by Levenberg-Marquardt approach in feedforward neural networks model. *Polymers*, 11(11). <https://doi.org/10.3390/polym11111853>
- [2] Aisien ET, Otuya IC, Aisien FA. Thermal and catalytic pyrolysis of waste polypropylene plastic using spent FCC catalyst. *Environ Technol Innov* 2021;22:101455. <https://doi.org/10.1016/J.ETI.2021.101455>.
- [3] Alabdrabalnabi, A., Gautam, R., & Mani Sarathy, S. (2022). Machine learning to predict pyro-char and pyro-oil yields from co-pyrolysis of biomass and plastics. *Fuel*, 328. <https://doi.org/10.1016/j.fuel.2022.125303>
- [4] Al-Yaari, M., & Dubdub, I. (2020). Application of artificial neural networks to predict the catalytic pyrolysis of HDPE using non-isothermal TGA data. *Polymers*, 12(8). <https://doi.org/10.3390/polym12081813>
- [5] Anuar Sharuddin SD, Abnisa F, Wan Daud WMA, Aroua MK. Energy recovery from pyrolysis of plastic waste: Study on non-recycled plastics (NRP) data as the real measure of plastic waste. *Energy Convers Manag* 2017;148:925–34. <https://doi.org/10.1016/j.enconman.2017.06.046>.
- [6] Areeprasert C, Asingsamanunt J, Srisawat S, Kaharn J, Inseemeeesak B, Phasee P, et al. Municipal Plastic Waste Composition Study at Transfer Station of Bangkok and Possibility of its Energy Recovery by Pyrolysis. *Energy Procedia* 2017;107:222–6. <https://doi.org/10.1016/j.egypro.2016.12.132>.

- [7] Ascher, S., Watson, I. and You, S. (2022) Machine learning methods for modelling the gasification and pyrolysis of biomass and waste. Renewable and Sustainable Energy Reviews, 155, 111902. <https://doi.org/10.1016/j.rser.2021.111902>
- [8] Bhattacharya M, Basak T. A review on the susceptor assisted microwave processing of materials. Energy 2016;97:306–38. <https://doi.org/10.1016/j.energy.2015.11.034>.
- [9] Cascone S, Gagliano A. Recycled agricultural plastic waste as green roof drainage layer within the perspective of ecological transition for the built environment. J Clean Prod 2022;380:135032. <https://doi.org/10.1016/j.jclepro.2022.135032>.
- [10] Chen W, Fu X, Chan W-P, Veksha A, Lisak G. Carbon nanosheet-carbon nanocage encapsulated Cu composite from chemical vapor deposition of real-world plastic waste for tailored CO₂ conversion to various products. Appl Mater Today 2021;25:101207. <https://doi.org/10.1016/j.apmt.2021.101207>.
- [11] Cheng Y, Ekici E, Yildiz G, Yang Y, Coward B, Wang J. Applied machine learning for prediction of waste plastic pyrolysis towards valuable fuel and chemicals production. J Anal Appl Pyrol 2023;169:105857.
- [12] Choudhary R, Mukhija A, Sharma S, Choudhary R, Chand A, Dewangan AK, et al. Energy-saving COVID–19 biomedical plastic waste treatment using the thermal - Catalytic pyrolysis. Energy 2023;264:126096. <https://doi.org/10.1016/J.ENERGY.2022.126096>.
- [13] Chow CF, Wong WL, Chan CW, Chan CS. Converting inert plastic waste into energetic materials: A study on the light-accelerated decomposition of plastic waste

- with the Fenton reaction. Waste Manag 2018;75:174–80.
<https://doi.org/10.1016/J.WASMAN.2018.01.034>.
- [14] Chow C-F, Wong W-L, Chan C-S, Li Y, Tang Q, Gong C-B. Breakdown of plastic waste into economically valuable carbon resources: Rapid and effective chemical treatment of polyvinylchloride with the Fenton catalyst. Polym Degrad Stab 2017;146:34–41. <https://doi.org/10.1016/j.polymdegradstab.2017.09.016>.
- [15] Cudjoe D, Wang H. Plasma gasification versus incineration of plastic waste: Energy, economic and environmental analysis. Fuel Process Technol 2022;237:107470. <https://doi.org/10.1016/J.FUPROC.2022.107470>.
- [16] Das P, Tiwari P. Valorization of packaging plastic waste by slow pyrolysis. Resour Conserv Recycl 2018;128:69–77.
<https://doi.org/10.1016/J.RESCONREC.2017.09.025>.
- [17] Daskalopoulos E, Badr O, Probert SD. An integrated approach to municipal solid waste management. Resour Conserv Recycl 1998;24:33–50.
[https://doi.org/10.1016/S0921-3449\(98\)00031-7](https://doi.org/10.1016/S0921-3449(98)00031-7).
- [18] Dubdub, I.; Al-Yaari, M. Pyrolysis of Mixed Plastic Waste: II. Artificial Neural Networks Prediction and Sensitivity Analysis. Appl. Sci. 2021, 11, 8456.
<https://doi.org/10.3390/app11188456>
- [19] Fu Z, Lin S, Tian H, Hao Y, Wu B, Liu S, et al. A comprehensive emission inventory of hazardous air pollutants from municipal solid waste incineration in China. Sci Total Environ 2022;826:154212.
<https://doi.org/10.1016/j.scitotenv.2022.154212>.

- [20] Galati A, Scalenghe R. Plastic end-of-life alternatives, with a focus on the agricultural sector. *Curr Opin Chem Eng* 2021;32:100681. <https://doi.org/10.1016/j.coche.2021.100681>.
- [21] Grimaldos-Osorio N, Sordello F, Passananti M, González-Cobos J, Bonhommé A, Vernoux P, et al. From plastic-waste to H₂: A first approach to the electrochemical reforming of dissolved Poly(methyl methacrylate) particles. *Int J Hydrogen Energy* 2022. <https://doi.org/10.1016/j.ijhydene.2022.02.229>.
- [22] Han M, Zhu S, Xia C, Yang B. Photocatalytic upcycling of poly(ethylene terephthalate) plastic to high-value chemicals. *Appl Catal B Environ* 2022;316:121662. <https://doi.org/10.1016/j.apcatb.2022.121662>.
- [23] Hasanzadeh R, Mojaver P, Azdast T, Chitsaz A, Park CB. Low-emission and energetically efficient co-gasification of coal by incorporating plastic waste: A modeling study. *Chemosphere* 2022;299:134408. <https://doi.org/10.1016/J.CHEMOSPHERE.2022.134408>.
- [24] Hendrawati, Liandi AR, Solehah M, Setyono MH, Aziz I, Siregar YDI. Pyrolysis of PP and HDPE from plastic packaging waste into liquid hydrocarbons using natural zeolite Lampung as a catalyst. *Case Stud Chem Environ Eng* 2023;7:100290. <https://doi.org/10.1016/J.CSCEE.2022.100290>.
- [25] Hu Q, Tang Z, Yao D, Yang H, Shao J, Chen H. Thermal behavior, kinetics and gas evolution characteristics for the co-pyrolysis of real-world plastic and tyre wastes. *J Clean Prod* 2020;260:121102. <https://doi.org/10.1016/J.JCLEPRO.2020.121102>.

- [26] Jeswani H, Krüger C, Russ M, Horlacher M, Antony F, Hann S, et al. Life cycle environmental impacts of chemical recycling via pyrolysis of mixed plastic waste in comparison with mechanical recycling and energy recovery. *Sci Total Environ* 2021;769:144483. <https://doi.org/10.1016/j.scitotenv.2020.144483>.
- [27] Kajiwara N, Noma Y, Tamiya M, Teranishi T, Kato Y, Ito Y, et al. Destruction of decabromodiphenyl ether during incineration of plastic television housing waste at commercial-scale industrial waste incineration plants. *J Environ Chem Eng* 2021;9:105172. <https://doi.org/10.1016/j.jece.2021.105172>.
- [28] Karimi Estahbanati MR, Kong XY, Eslami A, Soo H Sen. Current Developments in the Chemical Upcycling of Waste Plastics Using Alternative Energy Sources. *ChemSusChem* 2021;14:4152–66. <https://doi.org/10.1002/cssc.202100874>.
- [29] Khalid MY, Arif ZU, Ahmed W, Arshad H. Recent trends in recycling and reusing techniques of different plastic polymers and their composite materials. *Sustain Mater Technol* 2022;31:e00382. <https://doi.org/10.1016/j.susmat.2021.e00382>.
- [30] Kim TG, Srivastava RR, Jun M, Kim M, Lee J. Hydrometallurgical recycling of surface-coated metals from automobile-discarded ABS plastic waste. *Waste Manag* 2018;80:414–22. <https://doi.org/10.1016/j.wasman.2018.09.037>
- [31] KLAIMY S, LAMONIER JF, CASETTA M, HEYMANS S, DUQUESNE S. Recycling of plastic waste using flash pyrolysis – Effect of mixture composition. *Polym Degrad Stab* 2021;187:109540. <https://doi.org/10.1016/J.POLYMDEGRADSTAB.2021.109540>.

- [32] Lam SS, Chase HA. A Review on Waste to Energy Processes Using Microwave Pyrolysis. *Thermochem Waste Treat Combust Gasification, Other Methodol* 2016;221–56. <https://doi.org/10.1201/b19938-16>.
- [33] Leal Filho W, Salvia AL, Minhas A, Paço A, Dias-Ferreira C. The COVID-19 pandemic and single-use plastic waste in households: A preliminary study. *Sci Total Environ* 2021;793:148571. <https://doi.org/10.1016/j.scitotenv.2021.148571>.
- [34] Li, J., Liu, T., Niroshika Palansooriya, K., Yu, D., Wan, G., Sun, L., Chang, S. X., & Wang, Y. (n.d.). Zeolite-catalytic pyrolysis of waste plastics: Machine learning prediction, interpretation, and optimization. <https://ssrn.com/abstract=4925583>
- [35] Liu F, Zhuang X, Du Z, Dan Y, Huang Y, Jiang L. Enhanced photocatalytic performance by polarizing ferroelectric KNbO₃ for degradation of plastic wastes under mild conditions. *Appl Catal B Environ* 2022;318:121897. <https://doi.org/10.1016/j.apcatb.2022.121897>.
- [36] Martínez-Narro G, Prasertcharoensuk P, Diaz-Silvarrey LS, Dixon L, Phan AN. Chemical recycling of mixed plastic waste via catalytic pyrolysis. *J Environ Chem Eng* 2022;10:108494. <https://doi.org/10.1016/J.JECE.2022.108494>.
- [37] Mayanti B, Helo P. Closed-loop supply chain potential of agricultural plastic waste: Economic and environmental assessment of bale wrap waste recycling in Finland. *Int J Prod Econ* 2022;244:108347. <https://doi.org/10.1016/j.ijpe.2021.108347>.
- [38] Meng X, Fan W, Wan Mahari WA, Ge S, Xia C, Wu F, et al. Production of threedimensional fiber needle-punching composites from denim waste for utilization as furniture materials. *J Clean Prod* 2021;281:125321. <https://doi.org/10.1016/j.jclepro.2020.125321>.

- [39] Moo JGS, Veksha A, Oh W-D, Giannis A, Udayanga WDC, Lin S-X, et al. Plastic derived carbon nanotubes for electrocatalytic oxygen reduction reaction: Effects of plastic feedstock and synthesis temperature. *Electrochem Commun* 2019;101:11–8. <https://doi.org/10.1016/j.elecom.2019.02.014>.
- [40] Motasemi F, Ani FN. A review on microwave-assisted production of biodiesel. *Renew Sustain Energy Rev* 2012;16:4719–33. <https://doi.org/10.1016/j.rser.2012.03.069>.
- [41] Ortiz D, Munoz M, Nieto-Sandoval J, Romera-Castillo C, de Pedro ZM, Casas JA. Insights into the degradation of microplastics by Fenton oxidation: From surface modification to mineralization. *Chemosphere* 2022;309:136809. <https://doi.org/10.1016/j.chemosphere.2022.136809>.
- [42] Paul K, Chattopadhyay S, Dutta A, Krishna AP, Ray S. A comprehensive optimization model for integrated solid waste management system: A case study. *Environ Eng Res* 2018;24:220–37. <https://doi.org/10.4491/eer.2018.132>.
- [43] Peng Y, Wang Y, Ke L, Dai L, Wu Q, Cobb K, et al. A review on catalytic pyrolysis of plastic wastes to high-value products. *Energy Convers Manag* 2022;254:115243. <https://doi.org/10.1016/j.enconman.2022.115243>
- [44] Putra PHM, Rozali S, Patah MFA, Idris A. A review of microwave pyrolysis as a sustainable plastic waste management technique. *J Environ Manage* 2022;303:114240. <https://doi.org/10.1016/j.jenvman.2021.114240>.
- [45] Raj H, Bhatt S, Laddha S, Rana S. Ultrasonication-aided photocatalytic degradation of polyethylene. *Mater Today Proc* 2023;72:500–6. <https://doi.org/10.1016/j.matpr.2022.10.104>.

- [46] Ramos, E.; Lopes, A.G.; Mendonça, F. Application of Machine Learning in Plastic Waste Detection and Classification: A Systematic Review. *Processes* 2024, 12, 1632. <https://doi.org/10.3390/pr12081632>
- [47] Risco Á, Sucunza D, González-Egido S. Chemical recovery of waste electrical and electronic equipment by microwave-assisted pyrolysis: A review. *J Anal Appl Pyrolysis* 2021;159. <https://doi.org/10.1016/j.jaap.2021.105323>.
- [48] Schmidt S, Laner D. The multidimensional effects of single-use and packaging plastic strategies on German household waste management. *Waste Manag* 2021;131:187–200. <https://doi.org/10.1016/j.wasman.2021.06.003>.
- [49] scikit-learn documentation, "StandardScaler: Centering and Scaling Features," 2024
- [50] Shufian A, Mahin MSR, Islam R. Electronics and waste material recycler energyefficient electric furnace system. *Clean Eng Technol* 2022;6:100407. <https://doi.org/10.1016/j.clet.2022.100407>.
- [51] Siddique IJ, Salema AA, Antunes E, Vinu R. Technical challenges in scaling up the microwave technology for biomass processing. *Renew Sustain Energy Rev* 2022;153:111767. <https://doi.org/10.1016/j.rser.2021.111767>.
- [52] Singh RK, Ruj B, Sadhukhan AK, Gupta P. Impact of fast and slow pyrolysis on the degradation of mixed plastic waste: Product yield analysis and their characterization. *J Energy Inst* 2019;92:1647–57. <https://doi.org/10.1016/J.JOEI.2019.01.009>.

- [53] Su K, Liu H, Zhang C, Wang F. Photocatalytic conversion of waste plastics to low carbon number organic products. *Chinese J Catal* 2022;43:589–94. [https://doi.org/10.1016/S1872-2067\(21\)63885-8](https://doi.org/10.1016/S1872-2067(21)63885-8).
- [54] Tokmurzin D, Nam JY, Lee TR, Park SJ, Nam H, Yoon SJ, et al. High temperature flash pyrolysis characteristics of waste plastics (SRF) in a bubbling fluidized bed: Effect of temperature and pelletizing. *Fuel* 2022;326:125022. <https://doi.org/10.1016/J.FUEL.2022.125022>.
- [55] Veksha A, Yin K, Moo JGS, Oh W-D, Ahamed A, Chen WQ, et al. Processing of flexible plastic packaging waste into pyrolysis oil and multi-walled carbon nanotubes for electrocatalytic oxygen reduction. *J Hazard Mater* 2020;387:121256. <https://doi.org/10.1016/j.jhazmat.2019.121256>.
- [56] Wang J, Jiang J, Wang X, Wang R, Wang K, Pang S, et al. Converting polycarbonate and polystyrene plastic wastes into aromatic hydrocarbons via catalytic fast copyrolysis. *J Hazard Mater* 2020;386:121970. <https://doi.org/10.1016/J.JHAZMAT.2019.121970>.
- [57] Wang J, Liu L, Li S, Chen Y, Wang T, Zhang Y. Mercury emissions and distribution in a waste incineration plant based on the 30B and Ontario Hydro methods. *J Clean Prod* 2021;328:129663. <https://doi.org/10.1016/j.jclepro.2021.129663>.
- [58] Wang Z, Burra KG, Lei T, Gupta AK. Co-pyrolysis of waste plastic and solid biomass for synergistic production of biofuels and chemicals-A review. *Prog Energy Combust Sci* 2021;84. <https://doi.org/10.1016/j.pecs.2020.100899>.

- [59] Xue W, Cao K, Li W. Municipal solid waste collection optimization in Singapore. *Appl Geogr* 2015;62:182–90. <https://doi.org/10.1016/j.apgeog.2015.04.002>.
- [60] Xu D, Zhang Z, He Z, Wang S. Machine learning-driven prediction and optimization of monoaromatic oil production from catalytic co-pyrolysis of biomass and plastic wastes. *Fuel* 2023;350:128819.
- [61] Yamamoto M, Eva SN. What activities reduce plastic waste the most? – The path to a circular economy for Japan’s manufacturing industry. *Waste Manag* 2022;151:205–13. <https://doi.org/10.1016/j.wasman.2022.07.041>.
- [62] Yang X, Peng Y, Wang Y, Jiang L, Zhang S, Wu Q, et al. Conversion of soybean soapstock into hydrocarbon fuel by microwave-assisted catalytic fast pyrolysis using MCM-41/HZSM-5 in a downdraft reactor. *Chem Eng Process - Process Intensif* 2020;156:108109. <https://doi.org/10.1016/J.CEP.2020.108109>.
- [63] Zhang X, Xu R, Liu Q, Kong G, Lei H, Ruan R, et al. Enhancing the activity of Zn, Fe, and Ni-embedded microporous biocarbon: Towards efficiently catalytic fast copyrolysis/gasification of lignocellulosic and plastic wastes. *Energy Convers Manag X* 2022;13:100176. <https://doi.org/10.1016/J.ECMX.2021.100176>.

Vitae

Name	:Temirlan Shakhmanov
Nationality	:Kazakh
Date of Birth	:23.07.1996
Email	:temirshakh96@gmail.com
Address	:215 Baiken Ashimov St., Kokshetau, Akmola, Kazakhstan
Academic Background	:B.Eng and Tech. in Chemical Technology of Organic Substances, June 2018, Satbaayev University, Kazakhstan M.Sc. in Chemical Engineering, June 2025, KFUPM, Saudi Arabia.
Reference	: Dr. Mohammed Fahad Al-Khater, mfkhater@kfupm.edu.sa