



Using machine learning to predict the production and quality of bio-oil from pyrolysis biomass

Alireza Shafizadeh¹ , Mortaza Aghbashlo^{2*} , Meisam Tabatabaei³ , Hossein Mobli⁴ ,
Mohammad Hossein Nadian⁵ 

1- Department of Mechanical Engineering of Agricultural Machinery, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran, Email: a.shafizadeh@ut.ac.ir

2- Corresponding author, Department of Mechanical Engineering of Agricultural Machinery, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran, Email: maghashlo@ut.ac.ir

3- Microbial Biotechnology Department, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran, Email: meisam.tabatabaei@umt.edu.my

4- Department of Mechanical Engineering of Agricultural Machinery, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran, Email: hmobli@ut.ac.ir

5- Brain Engineering Research Center, Institute for Research in Fundamental Sciences, Tehran, Iran, Email: fmnh11510@gmail.com

Article Info

Article type: Research Article

Article history:

Received: May. 25, 2023

Revised: March. 1, 2022

Accepted: Apr. 4, 2022

Published online: Spring 2023

Keywords:

Biomass,
Thermochemical conversion,
Pyrolysis,
Modeling,
Machine learning.

ABSTRACT

Reducing the reserves of fossil energy sources serves as a warning sign for humanity. On the other hand, the increasing consumption of fossil fuels has led to significant environmental problems, such as global warming. These issues make the replacement of renewable energy sources with fossil fuels inevitable. Among various renewable energy sources, biomass is a reliable and sustainable resource. Thermochemical conversions of biomass are a promising method for converting raw biomass into liquid (bio-oil), solid (bio-char), and gas (biogas) fuels suitable for modern life. As one of the most important thermochemical conversions for efficient bio-oil production, pyrolysis has received significant attention. However, pyrolysis requires advanced equipment, precise product quantity, and quality measurement, which can be challenging and costly. Therefore, modeling has been extensively researched to enhance the performance and efficiency of pyrolysis. In recent years, machine learning has gained considerable attention in pyrolysis modeling, particularly for yield optimization, real-time monitoring, and process control. In addition to conventional techniques like artificial neural networks that capture nonlinear correlations between input and output values, combined machine learning models have been of particular interest for modeling and optimizing complex problems more effectively. This study provides a comprehensive overview of the research conducted on the application of machine learning in pyrolysis process modeling and assesses the prospects of this technology. These machine learning models have provided R² between 0.26 in the weakest case and 0.99 in the best case for predicting bio-oil production. These values have been presented between 0.6 and 0.93 to predict the improvement of bio-oil quality modeling.

Cite this article: Shafizadeh, A., Abhbashlo, M., Tabatabaei, M., Mobli, H., Nadian, M. H. (2023) Using machine learning to predict the production and quality of bio-oil from pyrolysis biomass, *Iranian Journal of Biosystem Engineering*, 54 (1), 87-113. <https://doi.org/10.22059/ijbse.2023.359872.665510>

© The Author(s).

Publisher: University of Tehran Press.



DOI: <https://doi.org/10.22059/ijbse.2023.359872.665510>

استفاده از یادگیری ماشین برای پیش‌بینی تولید و کیفیت روغن زیستی از زیست‌توده به روش پیرولیز

علیرضا شفیق‌زاده^۱، مرتضی آغباشلو^{۲*}، میثم طباطبایی^۳، حسین مبلی^۴، محمد حسین نادیان^۵^۱ انرژ‌های تجدیدپذیر، گروه مهندسی مکانیک ماشین‌های کشاورزی، دانشکده فنی و مهندسی کشاورزی، دانشکده کشاورزی و منابع طبیعی،دانشگاه تهران، کرج، ایران، رایانامه: a.shafizadeh@ut.ac.ir^۲ نویسنده مسئول، مهندسی ماشین‌های کشاورزی، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران، رایانامه: maghbashlo@ut.ac.ir^۳ بیوتکنولوژی میکروبی، پژوهشکده بیوتکنولوژی کشاورزی ایران، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران، رایانامه:meisam.tabatabaei@umt.edu.my^۴ مهندسی ماشین‌های کشاورزی، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران، رایانامه: hmobli@ut.ac.ir^۵ پژوهشکده علوم‌شناختی، پژوهشگاه دانش‌های بنیادی، تهران، ایران، fmnh11510@gmail.com

چکیده

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۲/۳/۴

تاریخ بازنگری: ۱۴۰۲/۸/۲

تاریخ پذیرش: ۱۴۰۲/۸/۷

تاریخ انتشار: بهار

واژه‌های کلیدی:

پیرولیز،

تبدیل حرارتی - شیمیایی،

زیست‌توده،

مدل‌سازی،

یادگیری ماشین.

کاهش ذخیره منابع انرژی‌های فسیلی یک زنگ خطر برای بشر است. از طرف دیگر، مصرف روبه‌رشد سوخت‌های فسیلی مشکل‌های زیست‌محیطی بسیاری مانند گرمایش زمین را با خود به همراه داشته است. این موارد جایگزینی انرژی‌های تجدیدپذیر را اجتناب‌ناپذیر ساخته است. در میان انواع انرژی‌های تجدیدپذیر زیست‌توده یکی از منابع قابل اطمینان و پایدار است. تبدیل‌های حرارتی - شیمیایی زیست‌توده به‌عنوان یک روش امیدوارکننده جهت تبدیل زیست‌توده خام به سوخت در حالت‌های مایع (روغن زیستی)، جامد (کربن زیستی) و گاز (گاز زیستی) در نظر گرفته شده است. پیرولیز به‌عنوان یکی از مهم‌ترین تبدیل‌های حرارتی - شیمیایی برای تولید مؤثر روغن زیستی موردتوجه گسترده قرار گرفته است. با این حال، باتوجه به پیچیدگی و نیاز به تجهیزات پیشرفته این فرایندها، اندازه‌گیری مقدار محصول^۳‌های تولید شده و کیفیت آنها به دلیل زمان و هزینه‌بر بودن بسیار چالش‌برانگیز است؛ بنابراین مدل‌سازی به‌عنوان یک شیوه مؤثر برای به حداقل رساندن عملکرد و بهره‌وری پیرولیز موردتوجه گسترده قرار گرفته است. در میان روش^۳‌های مختلف مدل‌سازی، یادگیری ماشین در سال‌های اخیر بخصوص برای بهینه‌سازی فرایند پیرولیز پیش‌بینی بازده، پایش بلادرنگ و کنترل فرایند توجه زیادی را به خود جلب کرده است. براین اساس، علاوه بر روش‌های پایه همچون شبکه‌های عصبی مصنوعی (یادگیری همبستگی‌های غیرخطی بین مقادیر ورودی و خروجی)، مدل‌های هم آمیخته یادگیری ماشین که از مدل‌های رایج برای مدل‌سازی و بهینه‌سازی مسائل پیچیده بسیار بهتر عمل می‌کنند موردتوجه خاص قرار گرفته‌اند. این مطالعه به طور جامع به تحقیق‌های صورت‌گرفته در مورد کاربردهای یادگیری ماشین در مدل‌سازی فرایند پیرولیز و چشم‌انداز پیشروی این فناوری می‌پردازد. این مدل^۳‌های ماشین یادگیری برای پیش‌بینی تولید روغن زیستی ضریب تعیین بین ۰/۲۶ در ضعیف‌ترین حالت و ۰/۹۹ را در بهترین حالت ارائه داده‌اند. این مدل‌ها مقادیر بین ۰/۶ و ۰/۹۳ را برای پیش‌بینی ارتقای کیفیت روغن زیستی ارائه نموده‌اند.

استناد: شفیق‌زاده؛ علیرضا، آغباشلو؛ مرتضی، طباطبایی؛ میثم، مبلی؛ حسین، نادیان؛ محمدحسین، (۱۴۰۲) استفاده از یادگیری ماشین برای پیش‌بینی تولید

و کیفیت روغن زیستی از زیست‌توده به روش پیرولیز، مجله مهندسی بیوسیستم ایران، ۵۴ (۱)، ۱۱۳-۸۷

<https://doi.org/10.22059/ijbse.2023.359872.665510>

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijbse.2023.359872.665510>

مقدمه

کاهش ذخایر سوخت فسیلی و نگرانی‌های فزاینده زیست‌محیطی ناشی از مصرف آنها و افزایش قیمت سوخت‌های فسیلی نیاز به جایگزینی منابع انرژی تجاری سازگار با محیط زیست را ضروری نموده است (Surenbra et al., 2022; Yang et al., 2021). زیست‌سوخت‌ها به عنوان یک منبع تجدیدپذیر انرژی و یک جایگزین بالقوه برای سوخت‌های فسیلی می‌تواند مورد توجه قرار گیرد (Awasthi et al., 2022; Velvizhi et al., 2022). تغییرات حرارتی-شیمیایی یکی از روش‌های موثر برای تبدیل زیست‌توده به زیست‌سوخت است. این موضوع بدلیل انعطاف در پردازش مواد اولیه، زمان عملیات کمتر، بازگشت سرمایه مناسب، امکان بهره‌گیری از طیف وسیعی از مواد اولیه، بازده فرآیند بالا و چرخه دی‌اکسیدکربن خنثی توجه زیادی را به خود جلب کرده است (Chen et al., 2021; Seo et al., 2022). علاوه بر اینها، امکان تولید سوخت به فرم‌های قابل حمل مایع (روغن زیستی)، جامد (کربن زیستی) از یک طرف و پتانسیل تولید مستقیم برق و گرما از طرف دیگر، فرایندهای تبدیل حرارتی - شیمیایی را به یکی از مهم‌ترین منابع قابل‌انرژی تجدیدپذیر تبدیل کرده است (Campbell et al., 2018; Madadian et al., 2020). بطوریکه از روغن زیستی تولیدی به عنوان سوخت خودروها و هواپیماها و از گاز زیستی برای تولید برق در نیروگاه‌های سیکل ترکیبی و از کربن زیستی یا بیوپچار تولیدی برای ترسیب کربن در خاک، جذب آلاینده‌ها استفاده شده است (Wurzer & Mašek, 2021).

روغن زیستی محصول اصلی فرایند پیرولیز است. این ماده حاوی ترکیب‌های مختلف مانند آلکان‌ها، هیدروکربن‌های معطر، مشتقات فنل و مقادیر کمی از کتون‌ها، استرها، اترها، قندها، آمین‌ها و الکل‌ها است. علاوه بر روغن زیستی، بیوپچار و گازهای غیرقابل تقطیر (هیدروژن، متان، دی‌اکسیدکربن و منو اکسیدکربن) از دیگر محصولات جانبی این فرایند هستند. در فرایند پیرولیز پیوندهای زیست پلیمر موجود در زیست‌توده توسط گرما شکسته و به واحدهای کوچکتری تبدیل می‌شوند. پیرولیز در راکتورهایی با محدوده دمایی ۳۰۰ تا ۶۵۰ درجه سلسیوس و در شرایط بدون وجود اکسیژن و در فشار اتمسفر انجام می‌پذیرد (Tang et al., 2020). در فرایند پیرولیز هم پارامترهای عملیاتی و هم خصوصیات زیست‌توده بر روی خروجی فرایند اثرگذار است. مواد خام آلی متنوعی تاکنون تحت عملیات پیرولیز قرار گرفته است که از جمله آنها می‌توان به بقایای کلزا و آفتابگردان، باگاس نیشکر، پوسته نارگیل و حتی فاضلاب انسانی اشاره کرد (Carpenter et al., 2014).

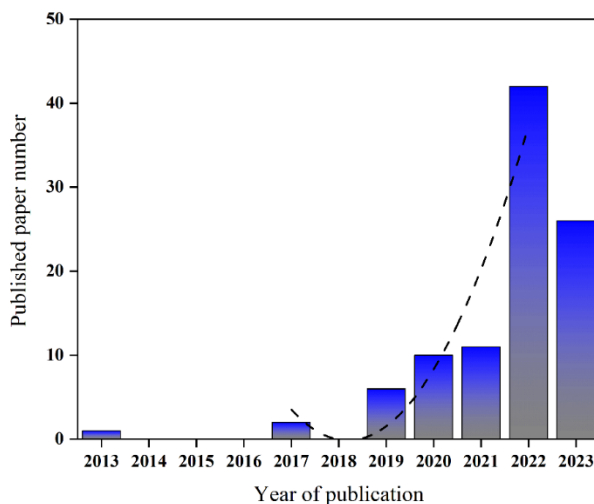
باتوجه به پیچیدگی‌های زیاد در فرایندهای تبدیل حرارتی-شیمیایی زیست‌توده و پارامترهای موثر بر آن، انجام مدل‌سازی برای درک مکانیسم‌های واقعی در طول فرایند، راه حل مناسبی است. مدل‌سازی به کاهش هزینه‌های آزمایش و تولید و همچنین صرفه جویی در وقت کمک شایانی می‌کند. از طرفی توسعه مدل‌های مختلف می‌تواند در طراحی فرایند پیرولیز، مطالعه اثرهای پارامترهای مختلف ورودی بر خروجی‌ها، پیش بینی مقدار کمیت و کیفیت محصول‌های تولیدی و شناسایی شرایط عملیاتی بهینه در شرایط صنعتی یا پایلوت مورد استفاده قرار گیرد (Basu, 2018). بر این اساس، توسعه مدل‌های مختلف برای مدل‌سازی فرایندهای شیمیایی حرارتی به طور کلی و فرایند پیرولیز به طور ویژه (بر مبنای داده‌های تجربی و یا با استفاده از مدل‌های ریاضی یا شبیه‌سازی)، تحقیقات قابل توجهی صورت گرفته است. در مواردی که دسترسی به داده‌های تجربی دشوار است، توسعه مدل‌های مبتنی بر مدل‌سازی ریاضی و آماری (Kumar et al., 2022) Sharma et al., 2022) معادله‌های سینتیکی، تعادل‌های شیمیایی و مدل‌های محاسباتی بر مبنای دینامیک سیالات می‌تواند در این زمینه یاری رسان باشد. اگرچه علیرغم توانایی‌های این روش‌ها، مواردی همچون وابستگی به جرم و انرژی و تکانه، فرایندهای محاسباتی سنگین و نیاز به تجهیزات رایانه‌ای پیشرفته برای حل محاسبات عددی، در نظر گرفتن پیش‌فرض‌های متعدد و ساده‌سازی‌های مختلف در حین انجام مدل‌سازی، محدودیت‌هایی را برای استفاده از این روش‌ها به وجود آورده است (Khan et al., 2023). به علاوه، به طور خاص در خصوص فرایندهای حرارتی-شیمیایی برای تولید زیست‌سوخت امکان پیش‌بینی کیفیت سوخت تولیدی در این روش‌های مدل‌سازی وجود ندارد (Shafizadeh et al., 2022).

در کنار روش‌های تئوری، تکنیک یادگیری ماشین^۱ که بر مبنای هوش مصنوعی^۲ توسعه یافته است، می‌تواند مدل‌سازی‌های دقیق و قابل‌انتکابی را بدون نیاز به انجام محاسبه‌های پیچیده و حل معادله‌های عددی ارائه دهد. هدف اصلی در یادگیری ماشین توسعه الگوریتمی است که به طور خودکار الگوها را از یک مجموعه داده مشخص (معمولاً خروجی داده‌های واقعی آزمایشگاهی یا صنعتی) استخراج کرده و

1 Machine learning

2 Artificial intelligence

پس از یادگیری این الگوها، موارد مورد نظر از توسعه مدل را بدون نیاز به دستورالعمل‌های پیچیده به صورت خروجی مدل و بسیار نزدیک به مقادیر واقعی ارائه دهد. از طرف دیگر روش‌های یادگیری ماشین در ترکیب با روش‌های تحلیل اهمیت پارامتر^۱ در مدل می‌تواند با ارزش‌گذاری میزان اثرگذاری و رده‌بندی آنها در حین مدل‌سازی، به شناسایی پارامترهای مهم و اثرگذار در شرایط واقعی به محقق کمک کند. با توجه به قابلیت‌های روش‌های یادگیری ماشین، مطالعه‌های مختلفی در زمینه استفاده از پتانسیل یادگیری ماشین در مدل‌سازی روش‌های شیمیایی-حرارتی برای تولید سوخت زیستی از زیست‌توده‌ها انجام شده است. به عنوان مثال یادگیری ماشین برای پیش‌بینی عملکرد و کیفیت نفت زیستی به‌دست آمده از روش مایع‌سازی گرمایی استفاده شده است (Shafizadeh et al., 2022). همچنین عملکرد و کیفیت روغن زیستی به‌دست آمده از پیرولیز پسماند فاضلاب نیز با روش‌های یادگیری ماشین مدل شده است (Shahbeik et al., 2022). در نمونه‌ای دیگر مدل‌سازی ارزیابی فرایند کربنی سازی گرمایی برای تولید بیوجار از زیست‌توده و بهینه‌سازی عملکرد محصول تولید شده برای مقاصد مختلف با استفاده از یادگیری ماشین ارائه شده است (Shafizadeh et al., 2023c). با توجه به انعطاف‌پذیری یادگیری ماشین، از آن برای مدل‌سازی فرایند اصلاح تولوئن به عنوان یک ترکیب پایه ناخواسته حاصل از فرایندهای گازی سازی یا پیرولیز استفاده شده است (Shafizadeh et al., 2023b). استفاده از یادگیری ماشین در مدل‌سازی فرایند پیرولیز در سال‌های اخیر رو به پیشرفت بوده است تعداد کل مقاله‌های چاپ شده بین سال‌های ۲۰۱۳ تا ۲۰۲۳ مرتبط با پیرولیز و ماشین یادگیری که در پایگاه داده اسکوپوس^۲ به چاپ رسیده‌اند در شکل ۱ نشان داده شده است. همانطور که در نمودار شماره یک مشخص است استفاده از ماشین یادگیری برای مدل‌سازی فرایند پیرولیز از سال ۲۰۱۹ روند صعودی گرفته است که نشان دهنده توجه پژوهشگران به پتانسیل‌های آن است.



شکل ۱. روند چاپ مقاله بین سال‌های ۲۰۱۳ تا ۲۰۲۳ که از ماشین یادگیری در فرایند پیرولیز استفاده کرده‌اند. شیوه جستجو در پایگاه داده به ترتیب ورود معادل انگلیسی "پیرولیز" و "زیست‌توده" و "ماشین یادگیری" بوده است.

باتوجه به اهمیت یادگیری ماشین در توسعه مدل‌های مربوط به روش‌های تبدیل حرارتی - شیمیایی زیست‌توده و روند روبه‌رشد استفاده از آن، این مقاله مروری به بررسی پژوهش‌های متأخر صورت‌گرفته در خصوص استفاده از یادگیری ماشین در تولید روغن زیستی از پیرولیز زیست‌توده و ارزیابی کیفی آن می‌پردازد. شایان‌ذکر است که در پایگاه‌های علمی و پژوهشی فارسی مقاله مروری که به طور کامل به بررسی پژوهش‌ها و پتانسیل استفاده از ماشین یادگیری در مدل‌سازی فرایندهای شیمیایی - حرارتی مانند پیرولیز بپردازد یافت نشد. این مقاله که به‌مرور جامع امکان استفاده از ماشین یادگیری برای مدل‌سازی فرایند پیرولیز می‌پردازد می‌تواند راهنمای مناسبی برای پژوهشگران جهت توسعه مدل‌های ماشین یادگیری برای مدل‌سازی در حوزه تبدیل‌های شیمیایی حرارتی زیست‌توده قرار گیرد. در این راستا ابتدا فرایند پیرولیز به طور جامع تعریف می‌شود. سپس، یادگیری ماشین تعریف و روش‌های مختلف استفاده شده در فرایند تولید و یا کیفیت‌سنجی روغن زیستی بررسی می‌گردد. در مرحله بعد تحقیق‌های صورت‌گرفته بر روی کاربرد یادگیری ماشین بر روی تولید روغن‌زیستی و بررسی کیفیت آن و نیز پارامترهای مؤثر بر آنها در مدل‌سازی پرداخته می‌شود. در بخش بعدی کاربرد تحلیل‌های اهمیت پارامتر در یادگیری ماشین مورد بررسی قرار می‌گیرد. در نهایت به مطالعات آینده و کاربردهای آتی و مشکلات پیشرو در این حوزه پرداخته

1 Feature importance analysis

2 Scopus

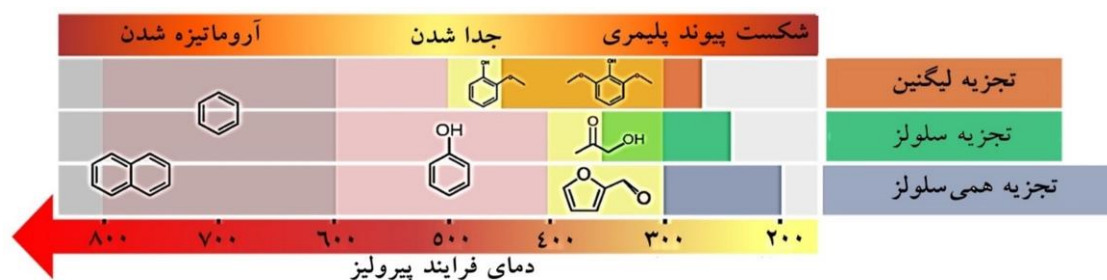
خواهد شد.

فرایند پیرولیز زیست‌توده

پیرولیز یکی از فرایندهای حرارتی - شیمیایی است که علاوه بر آنکه برای تولید انرژی از زیست‌توده استفاده می‌شود، به‌عنوان فرایند میانی برای فرایندهایی مانند گازی‌سازی نیز کاربرد دارد که در طول آن زیست‌توده تحت واکنش‌های اولیه و ثانویه شامل مکانیزم‌های انتقال جرم و حرارت قرار می‌گیرند و ساختار آن دچار شکست می‌گردد. در این فرایند واکنش‌های متعددی شامل آب‌زدایی، شکستن یا تشکیل مجدد پیوند بین پلیمرها، ایزومری شدن، آروماتیک شدن، کاهش دی‌اکسیدکربن به‌صورت موازی یا سری روی می‌دهد (Gautam et al., 2020).

پارامترهای متعددی بر روی پیرولیز زیست‌توده به‌عنوان یک فعل و انفعال پیچیده حرارتی و شیمیایی اثر گذار هستند. یکی از مهمترین پارامترها، نوع و گونه ماده خام مصرفی است. میزان سلولز، همی سلولز یا لیگنین زیست‌توده‌های لیگنوسلولوزیک در عملکرد روغن زیستی (مقادیر بالاتر سلولز و همی سلولز باعث عملکرد بیشتر روغن زیستی) و کیفیت روغن زیستی تولیدی (از نظر وجود ملکول‌های بزرگتر و ترکیب‌های پیچیده) اثرگذار است (D. Chen et al., 2022; Xiong et al., 2020). همچنین زیست‌توده‌های مختلف با داشتن ترکیب‌های عنصری (کربن، هیدروژن نیتروژن، سولفور و اکسیژن)، محتوای خاکستر، میزان رطوبت و مقدار حرارت ویژه بالاتر باعث ایجاد تفاوت در میزان عملکرد و کیفیت محصول‌های نهایی پیرولیز می‌شوند. وجود عناصر فلزی در بافت زیست‌توده (مانند پتاسیم، سدیم) در فرایند باعث عملکرد شبه کاتالیستی و تغییر کیفیت و کمیت محصول‌های فرایند می‌شوند (Di Blasi et al., 2009). مقدار بیشتر مواد فرار در زیست‌توده به دلیل آسان‌تر شدن واکنش‌های پیرولیز، باعث بهبود مقدار تولید روغن زیستی حین فرایند می‌شود (Abnisa et al., 2013). این در حالی است که مقدار بیشتر خاکستر در زیست‌توده باعث کاهش عملکرد تولید روغن زیستی و افزایش تولید بیوجار و گاز زیستی می‌شود (Venderbosch & Prins, 2010).

از پارامترهای عملیاتی موثر بر فرایند پیرولیز زیست‌توده می‌توان به دما، زمان واکنش، نرخ تغییرات دما و اندازه ذرات اشاره کرد. نقش اساسی دما فراهم کردن انرژی و گرمای لازم برای تخریب و شکستن پیوندهای ملکولی سلولز، همی سلولز و لیگنین زیست‌توده است. در دماهای کمتر از ۳۰۰ درجه سلسیوس معمولاً تخریب در هترو اتم‌های موجود در ساختار زیست‌توده رخ می‌دهد. در دماهای بالاتر (بیشتر از ۵۵۰ درجه سلسیوس) به دلیل تخریب زیاد ساختارهای زیست‌توده و رخداد واکنش‌های شکستن اولیه و ثانویه انواع مختلفی از ترکیب‌های ملکولی روغنی بوجود می‌آید. با به‌دست آوردن دمای بهینه با توجه به اثر آن بر واکنش‌های ثانویه، می‌توان مقادیر بهینه روغن زیستی و محصولات جانبی را کنترل نمود (Ding et al., 2018). دمای بهینه بر روی کیفیت روغن زیستی نیز موثر است. افزایش دمای واکنش نیز باعث کاهش محتوای اکسیژنی روغن زیستی تولیدی می‌گردد که در نتیجه آن محتوای بالاترین ارزش حرارتی روغن زیستی افزایش می‌یابد (Trinh et al., 2013). نرخ افزایش دما یکی دیگر از پارامترهای مهم در فرایند پیرولیز است که به‌ویژه بر روی ترکیب‌های روغن زیستی تولیدی اثرگذار است. نرخ‌های بالای افزایش حرارت باعث تخریب سریع پیوندهای زیست‌توده می‌شوند و باعث تولید مواد فرار بیشتری می‌شود. علاوه بر این نرخ‌های بالای افزایش حرارت باعث می‌شود که زمان محدودی برای واکنش‌های ثانویه مانند شکستن پیوندهای تار^۱ تولیدی و تشکیل مجدد پیوندها وجود داشته باشد (Akhtar & Saidina Amin, 2012). در مطالعه‌های مختلف رابطه سینرژی مثبت بین روغن زیستی و نرخ بالای افزایش دما گزارش شده است (Salehi et al., 2009). در شکل ۲ فرایند انجام پیرولیز در دماهای مختلف برای یک زیست‌توده لیگنوسلولوزیک حاوی سلولوز، همی سلولز و لیگنین و محصول‌های احتمالی تولیدی در هر دما نشان داده شده است.



شکل ۲. فرایندهای اصلی و محصول‌های احتمالی مختلف در نتیجه پیرولیز اجزا زیست‌توده در دماهای مختلف (Shafizadeh et al., 2023a).

مدت زمان انجام واکنش (زمان ماند بخارهای تولیدی تحت گرما) یکی دیگر از پارامترهای مهم در پیرولیز است. معمولاً زمان ماند از چند ثانیه تا چندین دقیقه توصیه می‌شود. اگرچه در زمان‌های ماند بسیار کوتاه فرصت شکستن همه پیوندها به دلیل دشواری انتقال حرارت در سطح ذره زیست‌توده، ممکن است فراهم نشود. همچنین باید در نظر داشت که زمان ماند ذرات زیست‌توده بایستی بیشتر از زمان ماند بخارهای حاصل از تخریب آنها باشد تا تبدیل حرارتی زیست‌توده کامل شود. بر مبنای این سه عامل مؤثر شامل سرعت و نرخ گرمایش، زمان ماند، و محدوده دمای عملیات، واکنش پیرولیز به سه زیر دسته کلی پیرولیز آهسته، سریع و بسیار سریع تقسیم می‌شود. پیرولیز آهسته که به نام کربن‌سازی نیز شناخته می‌شود، در دماهای پایین‌تر (مثلاً ۳۰۰ درجه سلسیوس) و زمان واکنش بسیار طولانی (از چند ساعت تا چند روز) رخ می‌دهد. پیرولیز سریع در نرخ افزایش گرمایش سریع (۱۰ تا ۲۰۰ درجه سلسیوس در دقیقه) و زمان ماند کوتاه (۰/۵ تا ۱۰ ثانیه، معمولاً ۲ ثانیه) و دماهای بالاتر (مثلاً بالاتر از ۷۰۰ درجه سلسیوس) اتفاق می‌افتد. این واکنش منجر به بازده بالای روغن زیستی ۵۰ تا ۷۰ درصد وزنی می‌شود. پیرولیز بسیار سریع با نرخ افزایش گرمایش بیشتر از پیرولیز سریع و زمان ماندن کوتاه‌تر (۰/۵ ثانیه) اتفاق می‌افتد. با این روش می‌توان بازده گاز زیستی بالاتری (۶۰-۵۰ درصد وزنی) به‌دست آورد (Foong et al., 2020; Kaczor et al., 2020).

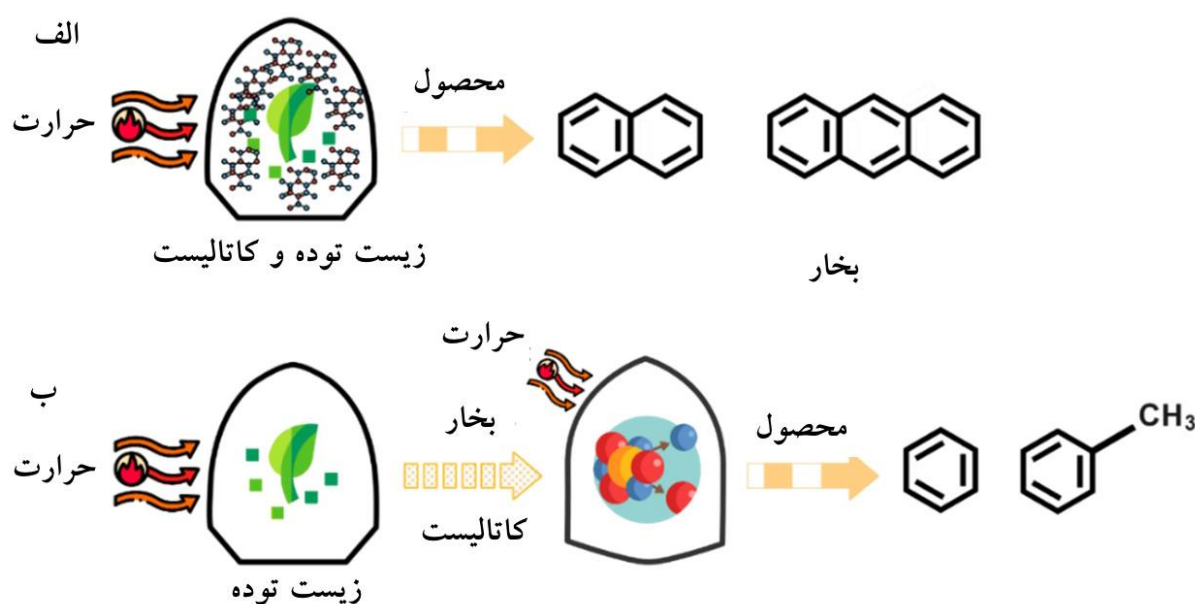
از دیگر پارامترهای مؤثر بر واکنش پیرولیز اندازه ذرات است. از آنجاکه ذرات زیست‌توده رسانای حرارتی خوبی نیستند (Suriapparao & Vinu, 2018)؛ بنابراین انتقال حرارت در حین فرایند پیرولیز به‌سختی انجام می‌شود و در نتیجه تغییر در اندازه ذرات و توزیع اندازه آنها می‌تواند به انتخاب زمان بهینه ماند کمک کند. به‌طور کلی ذرات کوچک‌تر به دلیل یکنواختی توزیع حرارت در آنها، نیاز به زمان‌های ماند کمتری دارند. در حالی که ذرات بزرگ‌تر با توجه به محدودیت بیشتر در انتقال حرارت، نیاز به انرژی فعال‌سازی بالاتری دارند و عموماً باعث کاهش عملکرد روغن زیستی می‌شوند (Haykiri-Acma, 2006; Pütün et al., 2007). محیط انجام واکنش پیرولیز بر روی شرایط و ترکیب‌های محصول‌های تولیدی اثرگذار است. اثرهای متقابل بین بخارهای ناشی از پیرولیز و ذرات جامد موجود در حین فرایند باعث سرعت گرفتن واکنش‌های ثانویه و افزایش مقدار کربن زیستی در حین انجام فرایند پیرولیز می‌شوند. برای حل این مشکل از تزریق گازهای خنثی مانند نیتروژن و آرگون برای خروج سریعتر بخارهای حاصل از پیرولیز استفاده می‌شود. بنابراین سرعت گازهای خنثی تزریق شده به فرایند بر روی زمان ماند بخارهای حاصل شده از پیرولیز در راکتور اثرگذار است. سرعت‌های کمتر گاز خنثی باعث دوام بیشتر این بخارها در راکتور شده و افزایش مقدار محصول‌های جامد را سبب می‌شود. این درحالی است که مقادیر بیشتر سرعت گاز خنثی باعث افزایش عملکرد محصول‌های گازی و کاهش محصول‌های جامد می‌شود (Varma et al., 2018).

روغن زیستی تولیدی با استفاده از روش پرولیز دارای کیفیت کافی، مناسب برای استفاده مستقیم برای کاربردهای انرژی نیست. این سوخت‌ها دارای ویسکوزیته نامناسب، محتوی اکسیژنی بالا و ترکیب‌های اکسیژن‌دار مختلف مانند الکل‌ها، الدهیدها، کتون‌ها، فنل‌ها و هیدروکربن‌های مختلف هستند که در نهایت این ترکیب‌های اکسیژن‌دار باعث کاهش ارزش حرارتی روغن زیستی می‌شوند (Shahbeik et al., 2023). استفاده از کاتالیزورها می‌تواند نقش مهمی در دستیابی به کیفیت بهتر روغن زیستی و تغییر در میزان عملکرد روغن زیستی از فرایند پیرولیز داشته باشند. با توجه به طبیعت فرایند پیرولیز که در واقع یک فرایند فاقد رطوبت و بر مبنای ماده خشک است از کاتالیست‌های ناهمگن یا همان هتروژن^۱ استفاده می‌شود (Li et al., 2020). این موضوع را می‌توان به برخی از ویژگی‌های مورد نظر کاتالیزورهای ناهمگن، مانند قابلیت بازیافت، قابلیت استفاده مجدد و پایداری نسبت داد (De Clercq et al., 2017). این کاتالیست‌های به‌زیردسته‌های کلی پایه معدنی و پایه فلزی و ترکیب آنها تقسیم‌بندی می‌شوند. کاتالیست‌هایی مانند انواع زئولیت، اکسیدهای فلزی، آلومینوسیلیکات، زئولیت‌هایی که بر روی آن کاتالیست فلزی قرار گرفته، چارچوب‌های فلزی-آلی^۲ از انواعی هستند که در فرایند پیرولیز مورد استفاده قرار می‌گیرند (Norouzi et al., 2021). در طی فرایند پیرولیز اندازه ذرات، ویژگی‌های ساختاری و فیزیکی و ویژگی‌های شیمیایی کاتالیست بر روی فرایند اثر گذار هستند. در فرایند پیرولیز استفاده از کاتالیست در ابعاد میکرو و یا نانو بر روی توزیع محصول‌ها و کیفیت آنها اثر گذار است (Shafizadeh et al., 2023a). همچنین از ویژگی‌های ساختاری مؤثر کاتالیست در فرایند پیرولیز می‌توان به سطح ویژه کاتالیست^۳، متوسط حجم منافذ^۴، متوسط قطر منافذ کاتالیست، درصد بلورینگی اشاره نمود (Zhou et al., 2023). از ویژگی‌های

1 Heterogeneous catalysts
2 Metal organic framework
3 BET surface area
4 Average pore volume

مهم شیمیایی می‌توان به مقدار سایت‌های اسیدی موجود بر سطح کاتالیست اشاره داشت؛ به نحوی که افزایش محتوی اسیدی کاتالیست می‌تواند باعث افزایش مقدار شکست پیوندهای زیست توده گردد (Liu et al., 2014).

فرایند پیرولیز زیست توده در حضور کاتالیست به دو شیوه درجا^۱ و برون جا^۲ انجام می‌شود (شکل ۳). در شیوه درجا کاتالیست و زیست توده در یک راکتور تحت تاثیر دمای واکنش قرار می‌گیرند درحالی که در شیوه برون جا، زیست توده در یک راکتور مجزا تحت دمای واکنش قرار می‌گیرد و سپس بخارهای پیرولیز حاصل از فرایند در راکتوری دیگر با کاتالیست تماس گرفته و کیفیت روغن زیستی تولیدی را بهبود می‌دهند. به طور کلی کاتالیزورها با کاهش انرژی فعال سازی و افزایش سرعت فعل و انفعالات مانند اکسیژن زدایی، شکستن پیوندهای ملکولی، آروماتیزاسیون و کتونیزاسیون باعث بهبود گزینش پذیری محصولات مطلوب، کاهش هزینه اقتصادی و بهبود عملکرد زیست محیطی فرایند پیرولیز می‌شوند (Al-Sabawi et al., 2012; Rahman et al., 2018). اضافه بر این موارد، انجام عملیات‌های پیش فراوری مانند اسیدشویی زیست می‌تواند با اثرگذاری بر روی محتوی اکسیژن آروغن زیستی، بر روی کیفیت نهایی آن در پیرولیز اثرگذار باشد (Kumar et al., 2009).



شکل ۳. فرایند پیرولیز زیست توده در حضور کاتالیست به شیوه الف) درون جا و ب) برون جا (Shafizadeh et al., 2023a).

روغن زیستی تولید شده از زیست توده برای کاربرد در مقاصد مختلف، نیاز به مطالعه جهت بررسی کیفیت و ارزیابی خصوصیات فیزیکی، شیمیایی و ترمودینامیکی دارد. کیفیت روغن زیستی را می‌توان براساس شاخص‌هایی مانند محتوی آبی، مقدار محتوی اسیدی و میزان اسیدی بودن، دانسیته، ویسکوزیته و بالاترین ارزش حرارتی تعیین کرد (Hassan et al., 2009). معیار "بالاترین ارزش حرارتی"^۳ یکی از مهمترین پارامترها در طراحی و عملیاتی کردن سیستم‌های انرژی از زیست توده است (Mateus et al., 2021). این معیار به صورت بیشینه مقدار انرژی قابل استحصال از زیست توده یا روغن زیستی تعریف می‌شود و به عنوان یک پارامتر مهم برای تعیین کیفیت سوخت کاربرد دارد. برای تعیین شاخص محتوی کالری یک سوخت زیستی نیاز به تجهیزات و صرف هزینه و زمان برای اندازه‌گیری آن است (Xing et al., 2019). همانند ارزیابی مقدار عملکرد پیرولیز زیست توده، بالاترین ارزش حرارتی نیز به پارامترهای مشخصات زیست توده و شرایط عملیاتی پیرولیز وابسته است (Leng et al., 2021). به عنوان مثال، مقدار بالاتر هیدروژن زیست توده منجر به بهبود بالاترین ارزش حرارتی روغن زیستی تولید شده می‌شود (Gani & Naruse, 2007). همچنین مقدار بالاتر محتوی اکسیژنی در زیست توده (مثلا مقدار بیشتر محتوی سلولوزی) باعث افزایش میزان محتوی اکسیژنی روغن زیستی شده و منجر به کاهش کیفیت و ارزش حرارتی آن می‌گردد (Trinh et al., 2013).

1 In-situ
2 Ex-situ
3 Higher heating value

یادگیری ماشین

یادگیری ماشین در زیرمجموعه هوش مصنوعی قرار می‌گیرد و به معنای توانایی یادگیری ماشین (رایانه) به یادگیری روابط پیچیده و کشف روابط بین پارامترهای مختلف در یک فرایند با استفاده از توسعه الگوریتم‌ها یا مدل‌های استاتیکی گفته می‌شود. مدل‌سازی انجام شده در یادگیری ماشین امکان انجام پیش‌بینی یا دسته‌بندی در خصوص یک فرایند مورد بررسی و براساس آموزش‌های ارائه شده به رایانه را فراهم می‌آورد (Samanpour et al., 2018). در ماشین یادگیری، فرایند یادگیری از طریق معرفی داده‌های تجربی قبلی یا به عبارت بهتر داده‌های آموزش و سپس استخراج الگوها و در نهایت ارائه نتایج این مدل‌ها به منظور رسیدن به یک هدف خاص (پیش‌بینی، خوشه‌بندی و دسته‌بندی) صورت می‌پذیرد. بدین صورت رایانه و مدل ماشین‌یادگیری با مواجه با مجموعه‌ای از داده‌ها که نتایج مشخصی را در ارتباط با هر نمونه داده خاص نشان می‌دهد تحت آموزش قرار می‌گیرد. به عبارت دیگر داده‌های مورد نیاز یادگیری ماشین جهت آموزش به صورت مجموعه‌ای از مثال‌ها تعریف می‌شوند. در این حالت هر مثال در برگرفته مجموعه‌ای از ویژگی‌ها است که تحت عنوان متغیرها از آنها یاد می‌شود. این متغیرها می‌توانند به فرم‌های مختلفی چون اعداد، نام، کدهای دودویی، ترتیبی و درجه‌بندی شده باشند (Liakos et al., 2018). روش‌های یادگیری ماشین با ساختن مجموعه‌ای از الگوریتم‌های خاص براساس متغیرها و سپس وزن‌دهی و تصحیح مداوم این وزن‌دهی‌ها می‌توانند فرایندهای پیچیده و روابط پنهان بین داده‌ها را شناسایی کنند و این عمل را تا رسیدن به مدل مناسب و بهینه که خروجی‌های قابل اطمینانی تولید کند را ادامه می‌دهند. این روش‌ها به‌خصوص برای پیش‌بینی نتایج حاصل از فرایندها با چندین ورودی مختلف (متغیرهای مستقل) جهت پیش‌بینی یا دسته‌بندی یک یا چند خروجی مرتبط با ورودی‌ها (متغیر وابسته) کاربرد فراوانی دارند (Somvanshi et al., 2016).

یادگیری ماشین به‌طور کلی به سه دسته کلی یادگیری بدون نظارت^۱، یادگیری با نظارت^۲ و یادگیری تقویتی^۳ تقسیم‌بندی می‌شود. رویکرد یادگیری بدون نظارت براساس یادگیری داده‌های بدون برچسب استوار است. هدف روش‌های یادگیری بدون نظارت آن است که زیرگروه‌ها یا خوشه‌های داده‌ها با الگوهای مشابه و کم تداخل که فاقد دخالت انسانی (کم یا زیاد) است را شناسایی کند. گاهی از این روش به عنوان تحلیل خوشه‌ای نام برده می‌شود. برخی از روش‌های یادگیری بدون نظارت معروف عبارتند از خوشه‌بندی مقدار میانی کا^۴، خوشه‌بندی سلسله مراتبی، تجزیه و تحلیل مؤلفه‌های اصلی و تشخیص ناهنجاری^۵ است (Harrison et al., 2021). در یادگیری نظارت شده، یادگیری براساس تحلیل متغیرهای ورودی برچسب دار برای پیش‌بینی یک یا چند خروجی برچسب دار به عنوان هدف مدل‌سازی صورت می‌پذیرد. بنابراین این روش یادگیری با تمرکز بر کشف یا تقریب رابطه بین متغیرهای ورودی و خروجی استوار است (Murali et al., 2020). آخرین مورد یادگیری تقویتی است که به عنوان یک رویکرد منحصر به فرد در آن فرآیند تصمیم‌گیری به صورت یک فرایند متوالی انجام می‌شود. در این روش امکان بهره‌گیری از برخی ویژگی‌های یادگیری نظارت شده و بدون نظارت به صورت اشتراکی وجود دارد. در این روش ماشین با استفاده از فرایند آزمایش و خطا عملیات یادگیری را انجام می‌دهد و به‌طور مداوم عملکرد خود را بر مبنای همان آزمون و خطاهای قبلی و تجربه‌های به‌دست آمده بهبود می‌دهد. این بهبود از طریق تطبیق پاسخ‌ها و عملکردها با شرایط و اعمال وزن‌دهی‌های جدید انجام می‌شود و باعث می‌گردد که مدل بهترین و بیشترین پیش‌بینی صحیح را بتواند ارائه کند (Chakraborty & Murphy, 2014).

روش یادگیری با نظارت در علوم مهندسی مانند پیش‌بینی فرایندهای حرارتی-شیمیایی، که مدل‌سازی بر مبنای به‌دست آوردن یک تابع هدف برای پیش‌بینی یک خروجی مشخص براساس پارامترهای ورودی صورت می‌پذیرد کاربرد بسیار فراوانی دارد. روش‌های یادگیری با نظارت را می‌توان به صورت ساده به ریاضی فرمول‌بندی نمود. در این شیوه مقدار متغیر خروجی وابسته (Y_i) با استفاده از مجموعه از متغیرهای مستقل ورودی (X_i) بر پایه شناسایی الگوریتم‌های بین آنها پیش‌بینی می‌شود. بنابراین مسئله اصلی پیدا کردن تابع $f(X)$ است که متغیرهای مستقل را به متغیرهای وابسته می‌رساند ($f: X \rightarrow Y, X \rightarrow f(X)$). برای ساخت و به‌دست آوردن یک پیش‌بینی کننده^۶ خوب از بین تمام توابع ساخته شده ($F = \{f: X \rightarrow Y\}$) نیاز به تعریف یک معیار عملکردی است. می‌توان مقدار تابع خطای (e) که وابسته به هر دو مقدار ورودی و خروجی است را تعریف کرد به نحوی که اگر خطای یک تابع (f) تولید شده از مقدار تابع دیگر (g) کمتر بود $e(f,)$

1 Unsupervised
2 Supervised
3 Reinforcement learning
4 k-means
5 Anomaly detection
6 Predictor

$(X, Y) < e(g, X, Y)$ تابع بهینه پیش‌بینی کننده به سادگی قابل تشخیص خواهد بود (Crisci et al., 2012). بدین ترتیب تابع پیش‌بینی کننده به صورت $Y=f(X)+e$ ارائه می‌گردد. بنابراین به طور کلی مدل بهینه f بایستی بتواند کمینه اختلاف بین خروجی مشاهده شده در شرایط واقعی حاصل از آزمایش $(y_{obs}=Y_i)$ و مقدار پیش‌بینی شده توسط تابع بهینه $(y_{pred} = f(X_i))$ برای تمام نمونه‌ها را ارائه کند. به اختلاف بین مقدار واقعی و مقدار پیش‌بینی شده تابع زیان گفته می‌شود. برای بیشتر مدل‌های رگرسیونی مورد استفاده در یادگیری ماشین با نظارت، توابع زیان شامل میانگین مربعات خطا^۱ (MSE) (Rashidi et al., 2023)، ریشه میانگین مربعات خطا^۲ (RMSE) و میانگین خطای مطلق^۳ (MAE) معمولاً برای ارزیابی و انتخاب مدل بهینه استفاده می‌شوند. علاوه بر توابع زیان، از معیار آماری ضریب تعیین نیز برای پیدا کردن بهترین مدل استفاده می‌شود. ضریب تعیین (R^2) بیانگر این موضوع است که وقتی یک متغیر وابسته حاصل از پیش‌بینی یادگیری ماشین به صورت خط رگرسیونی با نتایج حاصل از آزمایش قرار می‌گیرد چه میزان انحراف معیار را از خود نشان می‌دهد. ضریب تعیین معمولاً به صورت عددی بین صفر تا یک بیان می‌شود. هرچه قدر مقدار ضریب تعیین به عدد یک نزدیک‌تر باشد نشان دهنده توافق کامل داده‌ها با خط رگرسیون است و مدل تولید شده توسط یادگیری ماشین عملکرد بهتری را نشان می‌دهد، در حالی که مقدار ضریب تعیین برابر با صفر نشان دهنده هیچگونه توافق بین داده‌ها و خط رگرسیون نیست. معادله‌های توابع زیان اشاره شده و ضریب تعیین در معادله‌های (۱) تا (۴) ارائه شده است (Kasmuri et al., 2019; Shafizadeh et al., 2023c).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_p - y_a)^2}{\sum_{i=1}^n (y_a - \bar{y})^2} \quad (۱)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_p - y_a)^2 \quad (۲)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_p - y_a)^2} \quad (۳)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_a - y_p}{y_a} \right| \times 100 \quad (۴)$$

در این معادله‌ها y_p برابر با مقدار خروجی پیش‌بینی شده توسط مدل، y_a مقدار خروجی به‌دست آمده در شرایط واقعی، \bar{y} مقدار متوسط مقادیر خروجی در شرایط واقعی و n برابر با تعداد نمونه‌ها است.

مدل‌های پرکاربرد یادگیری ماشین با نظارت

مدل‌های متعددی در زیر شاخه یادگیری ماشین با نظارت وجود دارند که برخی در حوزه تولید و ارزیابی کیفیت روغن زیستی تولید شده از پیرولیز بیشتر به کار برده شده‌اند. از جمله این مدل‌ها می‌توان به ماشین یادگیری رگرسیون خطی چندگانه^۴، مدل رگرسیونی دسته‌بندی^۵، درخت تصمیم^۶ و جنگل تصادفی^۷، گرادیان تقویت شده درخت تصمیم^۸، الگوریتم تطبیقی آدابوست^۹، تقویت گرادیان شدید^{۱۰}، ماشین بردار پشتیبان^{۱۱}، مدل رگرسیون گوسین^{۱۲} و شبکه عصبی مصنوعی^{۱۳} اشاره کرد.

در ماشین یادگیری رگرسیون خطی چندگانه چندین مقادیر ورودی پیوسته به یک مقدار خروجی مرتبط و پیش‌بینی براین اساس انجام می‌شود. در این روش مقادیر مستقل به مقدار وابسته به وسیله رابطه خطی رگرسیونی تعریف می‌شود (Maulud & Abdulazeez, 2020).

- 1 Mean square error
- 2 Root mean square error
- 3 Mean absolute error
- 4 Multi linear regression
- 5 Bagging regression
- 6 Decision tree
- 7 Random forest
- 8 Gradient boosted decision tree
- 9 Adabtive boost
- 10 Extreme gradient boosting
- 11 Support vector machine
- 12 Gaussian process regression
- 13 Artificial neural network

رگرسیون دسته‌بندی شده زیرمجموعه روش یادگیری جمعی^۱ است که در آن از چندین مدل رگرسیونی با یک مدل پایه برای حل یک مسئله استفاده می‌شود. برای انجام پیش‌بینی موثر و ایجاد تنوع در مدل‌ها، هر مدل رگرسیونی روی زیرمجموعه‌های مختلف از مجموعه داده‌ها که به صورت تصادفی انتخاب و سپس جایگزین و به صورت جداگانه آموزش داده می‌شود. پس از آموزش، نتایج هر کدام از مدل‌ها با هم ترکیب می‌شود تا یک پاسخ نهایی براساس میانگین نتایج محاسبه هر کدام از آنها به دست آید (Xie et al., 2018).

تکنیک درخت تصمیم یک رویکرد بدون پارامتر برای توسعه مدل‌های پیش‌بینی در یادگیری ماشین با نظارت است. این مدل یادگیری ماشین شامل سه بخش اصلی است که شامل یک ریشه، شاخه‌ها و برگ‌ها یا یک گروه از گره‌های پایانی است. فرایند یادگیری از یک گره در بالای درخت (گره ریشه) شروع شده و سپس درخت از گره ریشه به شاخه‌های دودویی تقسیم می‌شود (رویکرد بالا به پایین). گره‌های جدیدی که ایجاد می‌شوند (گره‌های داخلی) داده تقسیم شده را دنبال می‌کنند. این تفکیک نماینده کلاس‌ها است و با بر اساس ویژگی‌های داده یا قوانین تصمیم‌گیری در طول فرایند تعریف می‌شود. این فرایند تا رسیدن به عمق مدل بهینه خود که شرایط توقف خاص (گره‌های برگ) پایانی هستند، ادامه می‌یابد. برای یافتن بهترین مدل‌سازی، درخت تصمیم با استفاده از مکانیزم جستجوگر^۲ به طور تکراری داده‌ها را به بخش‌های کوچک‌تری تقسیم می‌کند. در ضمن، این مدل یادگیری ماشین در کنار امکان مدل‌سازی پدیده‌های پیچیده، جزئیاتی درباره تفسیرپذیری مدل فراهم می‌کند (Shafizadeh et al., 2023c).

الگوریتم یادگیری جنگل تصادفی نمونه دیگری از شیوه‌های یادگیری بدون پارامتر و در زیر مجموعه مدل‌های جمعی قرار دارد که از مجموعه‌ای از درختان تصمیم با ساختار سلسله مراتبی تشکیل شده است. در این شیوه هر درخت با توجه به شرایط یا محدودیت‌هایی که مشخص می‌گردد، سلسله مراتب خود در مجموعه درخت‌های جنگل را مشخص می‌کند. بنابراین این قوانین عملکرد درخت از ریشه تا برگ را تعیین می‌کنند. این فرآیند با تولید بسیاری از نمونه‌ها به صورت تصادفی و با جایگزینی از مجموعه داده‌های آموزشی اصلی آغاز می‌شود. هر نمونه تصادفی با یک درخت رگرسیون مناسب برازش می‌شود. به این منظور برای هر گره در هر درخت، یک مجموعه کوچکی از متغیرهای مستقل ورودی از مجموعه کل متغیرهای مستقل به صورت تصادفی برای تقسیم دودویی در آن درخت تصمیم بخصوص در نظر گرفته می‌شود. معیار تقسیم در هر درخت بر اساس انتخاب متغیر ورودی با کمترین شاخص جینی است. بنابراین پیش‌بینی مقدار یک مشاهده با میانگین‌گیری بر روی مقادیر پیش‌بینی تمام درختان تصمیم صورت می‌گیرد (Wang et al., 2016). تکنیک جنگل تصادفی نیاز محاسباتی پایینی دارد. این الگوریتم در برابر داده‌های با پراکنش^۳ و پیش‌بینی بیش‌برازش^۴ مقاوم است. این الگوریتم همچنین خطای قابل اعتماد را نیز به دنبال دارد؛ بیشتر از همه، اهمیت متغیر از طریق این الگوریتم قابل توضیح است. علاوه بر این، الگوریتم جنگل تصادفی با مجموعه داده‌های بزرگ به طور کارآمد عمل می‌کند در حالی که به مقدار کمتری از فرآیند^۵ (تعداد درختان رگرسیون و تعداد متغیرهای ورودی در هر گره) نسبت به دیگر روش‌های یادگیری ماشینی نیاز دارد (M. Liu et al., 2014).

گرادینان تقویت شده درخت تصمیم یک روش یادگیری ماشینی جمعی دیگر است که از الگوریتم با جستجوی بسیار زیاد استفاده می‌کند. در این الگوریتم، به جای یک درخت، یک مدل بهبود یافته با بکارگیری مجموعه‌ای از درخت‌های تصمیم، خروجی‌های مطلوب را پیش‌بینی می‌کند. در این روش خروجی یک درخت تصمیم برای آموزش درخت تصمیم بعدی استفاده می‌شود. به عبارت بهتر، خروجی درخت‌های قبلی در فرآیند آموزش درخت‌های بعدی استفاده می‌شود. در مدل گرادینان تقویت شده درخت تصمیم، یادگیرنده‌های ضعیف با توجه به خطاهای حاصل در هر مرحله با هم ترکیب می‌شوند تا یک مدل پیش‌بینی قوی به عنوان مجموع درخت‌های ضعیف متوالی حاصل شود (Nie et al., 2021). الگوریتم کاهش گرادینان برای کمینه کردن خطا در مدل‌های متوالی استفاده می‌شود (Rzychoń et al., 2021). احتمال بیش‌برازش در الگوریتم درخت تصمیم گرادینان تقویت شده به دلیل تصادفی کردن در انتخاب مجموعه آموزشی، کمتر است. دو پارامتر اصلی این شیوه تعداد تکرار و نرخ یادگیری هستند. این پارامترها می‌توانند پیچیدگی مدل را به خوبی نشان دهند. افزایش تعداد تکرار و نرخ یادگیری می‌تواند مدل قوی‌تری را تولید کند و به جلوگیری از بیش‌برازش مدل کمک کند (Ghafariyan et al., 2022). تقویت گرادینان شدید الگوی ارتقا یافته الگوریتم یادگیری ماشین درخت تصمیم گرادینان تقویت شده است. با این تفاوت که این الگوریتم با اضافه کردن درخت‌های تصمیم جدید به مدل به صورت تکراری، هر کدام از خطاهایی که در درخت‌های تصمیم پیشین وجود داشته‌اند

1 Ensemble

2 Greedy

3 Noisy data

4 Overfitting

5 Hyperparameter

را اصلاح می‌کند. به عبارت دیگر، این روش از روش درخت تصمیم گرادیان تقویت شده ساده استفاده می‌کند و به صورت تدریجی بهترین مدل را تولید می‌کند. این شیوه از گرادیان آماری مرتبه اول و دوم برای بهینه سازی تابع زیان استفاده می‌کند. این الگوریتم علاوه بر حفظ خصوصیات مدل پایه خود مانند پیشگیری از بیش‌برازش، کاهش زمان محاسبه‌ها از تطبیق‌پذیری و قدرت بالایی در پیش‌بینی مسائل پیچیده برخوردار است (W. Zhang et al., 2021).

الگوریتم تقویتی سازگار شونده نیز یک الگوریتم جمعی دیگر است که از ترکیب کردن چندین دسته‌بندی ساده و با پیش‌بینی ضعیف استفاده و با بهبود مداوم آنها مدل نهایی را ارائه می‌دهد. در این روش به طور خاص دسته‌بندی‌ها به صورت تصادفی انتخاب نمی‌شوند. در طول فرایند آموزش به‌روزرسانی مجدد وزن‌های دسته‌های آموزش‌دیده بر اساس مقدار خطای آن‌ها در مرحله قبلی انجام می‌پذیرد. به نحوی که در هر مرحله، یک دسته‌بندی قوی‌تر جدید با استفاده از وزن‌های به‌روز شده آموزش‌داده شده پیشین ایجاد می‌شود و با استفاده از یک مجموع وزن‌دار از پیش‌بینی‌های آن به مجموعه اضافه می‌شود (Kim & Philen, 2011).

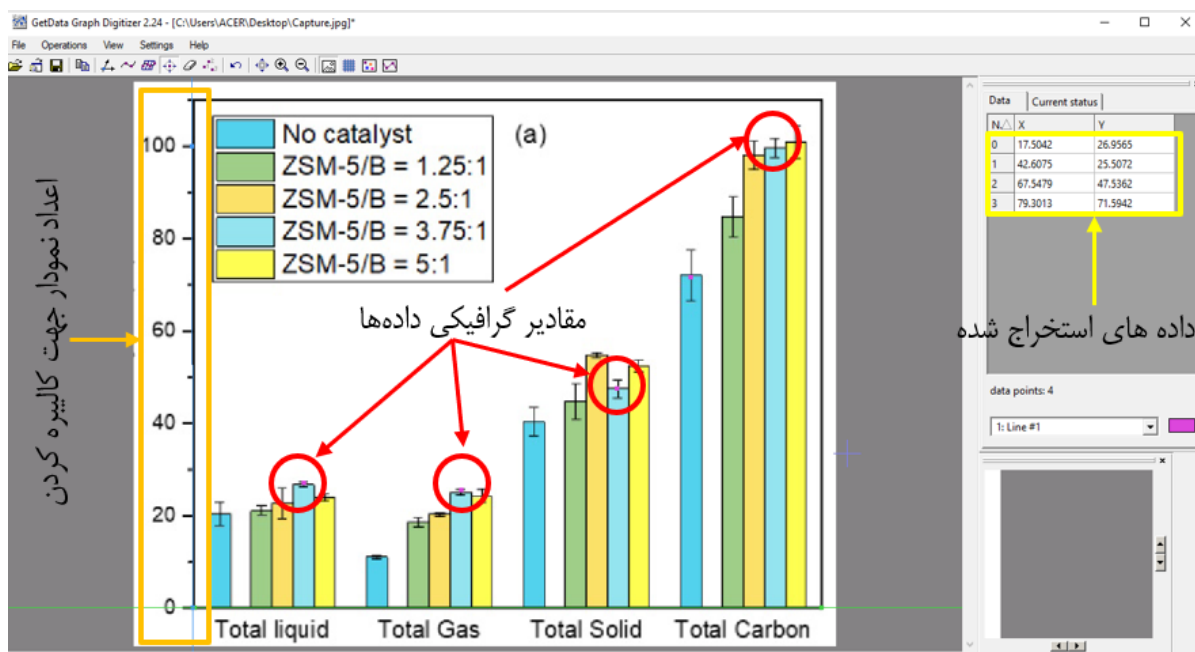
روش ماشین بردار پشتیبان، یک روش قدرتمند توسعه‌یافته بر مبنای روش‌های آماری است که برای حل مسائل مدل‌سازی خطی و غیرخطی در بسیاری از حوزه‌ها کاربرد دارد. ماشین بردار پشتیبان سعی می‌کند با استفاده از یک تابع هسته مناسب داده‌ها را به گروه‌های مختلف تقسیم کرده و یک یا چند خط (و یا صفحه) مجزاکننده در فضای بالا بعد یا بی‌نهایت بعد ساخته و در نهایت معادله رگرسیونی را شکل دهد. خط (صفحه) بهینه با افزایش حاشیه بین نزدیک‌ترین نقطه آموزش و آن خط انتخاب می‌شود. ماشین بردار پشتیبان به دلیل توانایی مدیریت نمونه‌های داده کم، قابلیت سازگاری بالا و عملکرد تعمیم‌پذیر خوب، یک مدل مناسب برای مسائل طبقه‌بندی و رگرسیون است (Raghavendra & Deka, 2014).

مدل رگرسیون گوسین یک روش دیگری ماشین یادگیری با نظارت است. این شیوه یک روش غیرپارامتری احتمالی است که از توابع کرنل^۱ برای افزایش قابلیت مدل‌سازی فرایندهای غیرخطی استفاده می‌کند. این مدل از روش کریجینگ^۲ که یک شیوه درون‌یابی بر اساس ماتریس‌های انحراف معیار است، برای تجزیه و تحلیل‌های رگرسیونی استفاده می‌کند. کریجینگ مقادیر خروجی یک پیش‌بینی جدید بر مبنای داده شده را با استفاده از مقادیر خروجی مشاهده شده در نقاط قدیمی وارد شده در ورودی و خروجی‌ها، شبیه‌سازی می‌کند. مدل رگرسیون گوسین قادر است منحنی‌هایی با هر شکل و درجه غیرخطی را تنظیم کند. به علاوه، این مدل می‌تواند با متغیرهای چندبعدی سازگاری پیدا کند در حالی که محاسبات پیش‌بینی در این روش تقریباً ساده هستند و نیاز به توان محاسباتی بالایی ندارند (Shafizadeh et al., 2022). شبکه عصبی مصنوعی، یک روش یادگیری ماشین است که از ایده شبیه‌سازی مغز انسان شکل گرفته است. در مقابل رویکرد رگرسیون سنتی، شبکه عصبی مصنوعی قادر به مدل‌سازی روابط غیرخطی پیچیده است. همچنین، شبکه عصبی مصنوعی دارای تحمل خطا عالی است و با پردازش موازی سریع و قابل توسعه بسیار خوب عمل می‌کند. حداقل از سه لایه شامل یعنی یک لایه ورودی، یک لایه پنهان و یک لایه خروجی تشکیل شده است. هر لایه از چندین نورون تشکیل شده که این نورون‌ها با نورون‌ها در لایه بعدی با مقادیر وزنی خاص متصل می‌شود. این وزن‌ها با استفاده از یک تکنیک یادگیری نظارت شده به نام پس‌خور یا پیش‌خور تعیین می‌شوند. مقدار پیش‌بینی در خروجی بر اساس این وزن‌دهی‌ها در لایه پنهان و تصحیح مداوم آنها تا رسیدن به نتیجه مطلوب ادامه می‌یابد. این روش به خاطر قابلیت تطبیق‌پذیری خوب امکان ارائه راه‌حل‌های تقریبی خوب برای مسائل بسیار پیچیده را دارد (Shahbeik et al., 2022).

کاربرد یادگیری ماشین در پیش‌بینی تولید روغن زیستی و کیفیت آن

استفاده از روش‌های کارآمد ماشین یادگیری برای کاربردهای مختلف در مدل‌سازی فرایند پیرولیز زیست‌توده از سابقه چندانی برخوردار نیست. اولین تحقیق‌ها در این زمینه در ۲۰۱۳ ارائه شده است. در طول این مدت استفاده از شبکه‌های عصبی مصنوعی با توجه به قابلیت آنها به یادگیری روابط غیرخطی با اقبال فراوانی روبرو بوده‌اند. علاوه بر روش‌های مرتبط با هوش مصنوعی، روش‌های مبتنی بر درخت تصمیم یا روش‌های پیچیده‌تر مبتنی بر روش هم آمیخته یادگیری ماشین برای پیش‌بینی فرایند پیرولیز زیست‌توده مورد کاربرد بسیاری بوده است. از روش‌های پرکاربرد در این حوزه می‌توان به جنگل تصادفی اشاره کرد. از دیگر روش‌های پر کاربرد یادگیری ماشین می‌توان به زیرمجموعه‌های دسته‌بندی داده‌ها مانند شیوه ماشین بردار پشتیبان اشاره کرد که عملکرد بسیار خوبی را در مواردی از خود نشان داده است.

برای مدل سازی تولید روغن زیستی توسط شیوه های یادگیری ماشین اولین قدم تشکیل بانک داده مناسب است. به طور کلی دو شیوه برای تشکیل بانک داده وجود دارد یکی تولید داده با انجام آزمایش های متعدد که روشی زمان بر و پرهزینه است و دیگری استخراج داده از مقاله ها و پژوهش های چاپ شده که این روش می تواند حجم قابل توجهی از داده های مورد نیاز برای مدل سازی را در اختیار قرار دهد؛ بنابراین با تعیین متغیرهای مستقل ورودی و متغیرهای وابسته خروجی جهت مدل سازی استخراج داده از مقاله ها صورت می پذیرد. برای این منظور معمولاً بانک داده به صورت یک فایل اکسل تهیه می گردد که این فایل شامل ستون هایی مجزا است که هر ستون بر اساس نام متغیرهای ورودی و خروجی عنوان گذاری شده است. داده های هر ستون مورد نیاز برای استفاده در مدل سازی یک فرایند مشخص از درون پژوهش های معتبر چاپ شده استخراج می گردد. این داده ها به شیوه های مختلفی در پژوهش ها گزارش شده اند. بسیاری از داده ها به صورت عددی در جدول یا در متن، یا به صورت نمودارهای گرافیکی (ستونی یا دایره ای) گزارش می شوند؛ بنابراین استخراج هر کدام از آنها به فرایندی مناسب خود نیازمند است. معمولاً داده های مربوط به هر پارامتر ورودی یا خروجی که به صورت عدد در جدول یا متن مقاله گزارش شده اند به همان صورت (در صورت نیاز با یکسان سازی یکاها) در فایل نهایی در ستون مربوط به خود وارد می شوند. برای استخراج داده هایی که به صورت گراف و نمودار ستونی در پژوهش ها گزارش شده اند فرایند استخراج داده در طی چهار مرحله مجزا صورت می پذیرد. در مرحله اول یک تصویر با کیفیت از نمودار گزارش شده در پژوهش تهیه می شود. سپس این تصویر با فرمت مناسب (jpg) به یک نرم افزار دیجیتال سازی نمودار مانند (مثلاً نرم افزار getdata فراخوانی می گردد (شکل ۴). در این نرم افزار پیکسل های تصویر گرافیکی با اعداد نمودار ارائه شده در تصویر کالیبره شده و داده ها به صورت مختصات (X,Y) استخراج می گردند. بایستی دقت داشت که کدام محور نمودار برای استخراج داده عددی دارای پتانسیل است و اگر محوری از نمودار به صورت عدد گزارش نشده بود (مثلاً مقدار عملکرد محصول های مختلف در محور X، بایستی در اکسل مورد نظر بجای مقدار محور X نام محصول درج و مقدار عددی در محور دیگر در مقابل یا در ستون مربوطه در فایل اکسل مربوطه درج گردد.



شکل ۴. نمونه ایی از تبدیل اطلاعات گرافیکی یک نمودار به داده های دیجیتال (عدد) در نرم افزار getdata.

مرحله بعدی پس از ایجاد بانک داده ارزیابی داده های جمع آوری شده از نظر اعتبار و صحت استفاده برای مدل سازی است. این موضوع بخصوص در مواردی که بانک داده از مجموعه پژوهش های به چاپ رسیده ایجاد شده باشد دارای اهمیت بیشتری است. در این مرحله داده های پرت بایستی از بانک داده خارج گردند.

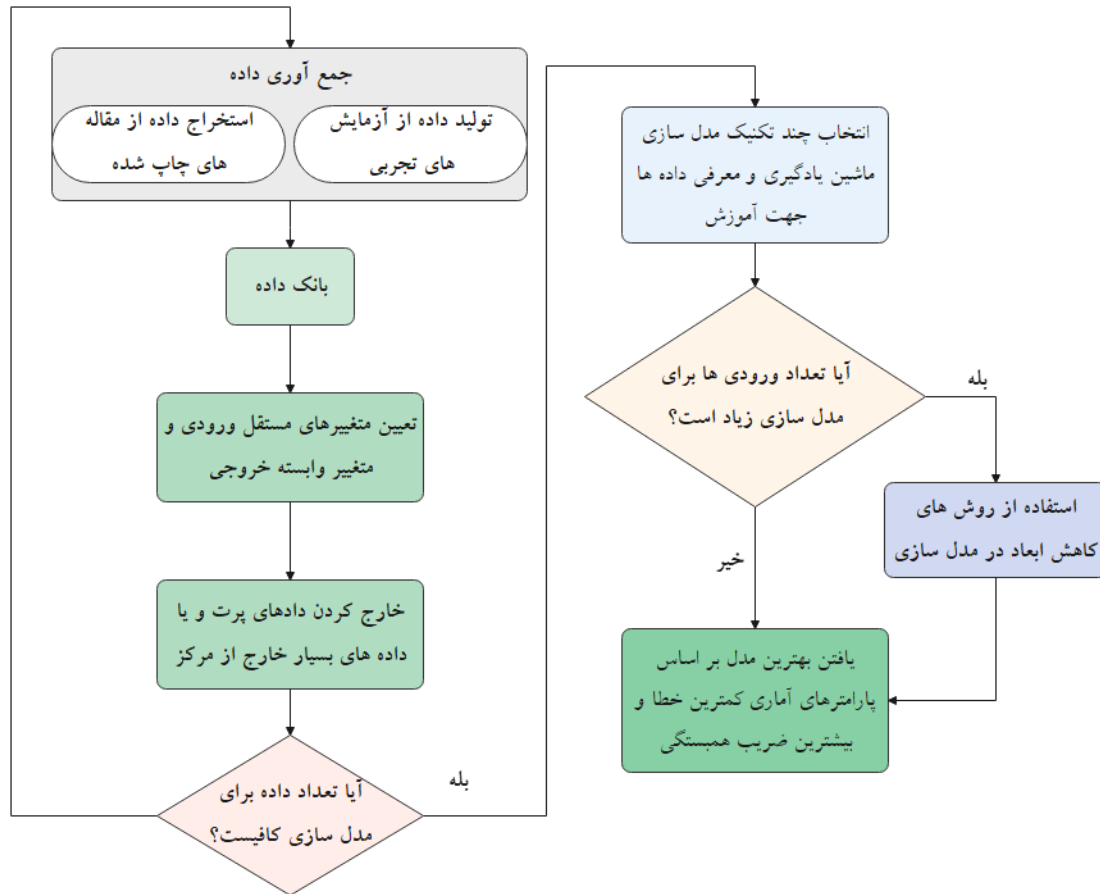
برای خارج کردن داده های پرت از بانک داده می توان به دو صورت اعتماد به تجربه متخصص در زمینه مدنظر از لحاظ اعتبارسنجی داده ها و همچنین بر اساس روش های آماری اقدام نمود. در مورد خارج کردن داده های پرت بر مبنای دانش متخصص، داده ها از لحاظ علمی مورد بررسی قرار می گیرند و منطقی بودن آنها ارزیابی می گردد. علاوه بر این مواردی همچون فاصله داده های یک پارامتر از هم و

همچنین وجود تعداد داده مناسب بین دو مقدار اصلی متوالی باید مدنظر قرار گیرد. مثلاً در مورد دمای پیرولیز دماهای بین ۲۵۰ تا ۸۰۰ قابل قبول هستند بنابراین باید بین این دو داده تعداد مناسبی از داده‌های دمایی با فواصل مناسب مثلاً ۲۰ درجه‌ای موجود باشد تا مدل ماشین یادگیری در این بازه قابل اتکا باشد. همچنین اگر در بانک داده مثلاً ۲ عدد داده از مقدار دماهای ۱۱۰۰ و ۱۲۰۰ درجه وجود داشت این داده‌ها به دلیل فاصله زیاد از رنج معمول کاربرد و همچنین خلاء وجود تعداد داده با فواصل مناسب از بازه، بایستی حذف گردند. این امر به آن دلیل است که عملاً امکان مدل‌سازی بین فاصله ۸۰۰ تا ۱۲۰۰ درجه و همچنین مدل‌سازی خود دماهای ۱۱۰۰ و ۱۲۰۰ به دلیل تعداد کم گزارش داده وجود نخواهد داشت. از طرف دیگر ممکن است برخی مقادیر از نظر متخصص در دامنه قابل قبول قرار داشته باشند اما با بررسی بانک داده با برخی روش‌ها و توزیع‌های آماری در دسترس به صورت داده‌های پرت تشخیص داده شوند که این موارد نیز بهتر است از فرایند مدل‌سازی حذف گردند. این روش‌ها می‌تواند شامل برخی روش‌های آماری مانند روش گرابز^۱ باشد که در آن مقدار فاصله بین داده و متوسط داده‌ها به عنوان معیاری برای تشخیص پرت بودن داده استفاده می‌گردد. علاوه بر این، روش‌های دسته‌بندی نظارت نشده^۲ یا برخی روش‌های نظارت شده که در آن داده‌ها نرمال و غیرنرمال برچسب گذاری می‌گردند نیز برای تشخیص داده‌های پرت استفاده می‌گردند (Hodge & Austin, 2004).

علاوه بر خارج نمودن داده‌های پرت، برای آنکه مدل بخوبی آموزش ببیند، وجود تعداد مناسب داده در بین بیشینه و کمینه همه متغیرها از اولویت برخوردار است. نهایتاً پس از خارج کردن داده‌های پرت و یا داده‌ها با فاصله بسیار زیاد از مرکز اصلی داده‌ها که فاقد داده‌ها واسط میانی جهت آموزش هستند، باید تعداد مناسبی داده برای آموزش ماشین یادگیری وجود داشته باشد. البته هیچ قانون مدونی برای اینکه چه تعداد داده برای مدل‌سازی نظارت شده در علوم مهندسی کافی است وجود ندارد اما طبق یک روال حداقل تعداد ردیف داده براساس تعداد متغیرهای مستقل قابل تخمین است به نحوی که حداقل ردیف داده برای آموزش مدل ۱۰ برابر تعداد متغیرهای مستقل خواهد بود. به عنوان مثال اگر در یک مدل‌سازی تعداد متغیرهای مستقل ۹ باشد تعداد ردیف داده مورد نیاز برای مدل‌سازی متغیر خروجی ۹۰ ردیف داده خواهد بود (Zeng et al., 2022). بایستی توجه داشت که معمولاً ۸۰ درصد داده‌های بانک داده برای آموزش و ۲۰ درصد باقیمانده برای تست مدل استفاده می‌شوند. پس از این مرحله یک یا چند مدل یادگیری ماشین با توجه به ماهیت داده‌ها انتخاب و برای آموزش داده‌ها به آن‌ها معرفی می‌شود (Y. Yang et al., 2023).

بایستی توجه داشت در طول فرایند مدل‌سازی ممکن است برخی مدل‌ها دچار بیش برآزش^۳ شوند. بیش برآزش در یادگیری ماشین به حالتی اطلاق می‌شود که در یک مدل جزئیات و نویز داده‌ها نیز در فرایند آموزش وارد می‌شوند که این موضوع بر عملکرد مدل وقتی در برابر داده‌های جدید و یا نادیده قرار می‌گیرد، تأثیر منفی می‌گذارد، به نحوی که مدل بر روی داده‌های آموزش بیش برآزش می‌شود درحالی‌که در مورد ارائه خروجی در مورد داده‌های نادیده ضعیف برخورد می‌کند. عوامل مؤثری بر ایجاد بیش برآزش عدم وجود داده کافی در بانک داده، پیچیدگی بیش از اندازه مسئله، وارد نمودن ویژگی‌های نامربوط در ورودی‌های مدل، وجود داده‌ها با فواصل زیاد، عدم انجام عملیات تطبیق بر روی پارامترها با ضریب تأثیر بسیار زیاد و عدم اعتبارسنجی متقابل است. از نشانه‌های بیش برآزش می‌توان به دقت آموزش بالا اما دقت اعتبارسنجی پایین در مرحله تست، پیش‌بینی‌های ناپایدار و یا نامنظم و تعمیم ضعیف بر روی داده‌های دیده نشده اشاره نمود (Ying, 2019). در مواردی که تعداد متغیرهای ورودی بسیار زیاد می‌شود و امکان بیش برآزش وجود داشته باشد استفاده از روش‌های کاهش ابعاد مانند تجزیه و تحلیل مولفه‌های اصلی^۴ در حین مدل‌سازی قابل استفاده خواهد بود. با استفاده از روش‌های کاهش ابعاد پارامترهای مستقل ورودی که بیشترین اثر در خروجی‌ها دارند شناسایی می‌شوند و مدل‌سازی نهایی تنها بر همان مبنا انجام خواهد پذیرفت. در نهایت مدل‌های آموزش دیده بر مبنای شاخص‌های آماری ارزیابی می‌شوند و بهترین مدل آموزش دیده برای کاربردهای دیگر انتخاب خواهد شد. در شکل ۵ روند کلی انجام مدل‌سازی با روش‌های ماشین یادگیری با استفاده از یک بانک داده ایجاد شده به صورت طرح‌واره ارائه شده است.

1 Grubbs method
 2 Unsupervised clustering
 3 Overfitting
 4 Principle component analysis



شکل ۵. روند مدل سازی و گام های اصلی با استفاده از تکنیک های ماشین یادگیری

برای مدل سازی فرایند پیرولیز در مطالعات مختلف ورودی های مستقل متنوعی در نظر گرفته شده است. از نظر ترکیب زیست توده مواردی همچون مقدار محتوی کربنی، هیدروژن، اکسیژن، نیتروژن، سولفور زیست توده و همچنین مقدار محتوی رطوبتی، میزان خاکستر و مقدار کربن تثبیت شده آن به عنوان پارامترهای مؤثر بر فرایند پیرولیز در ورودی های مستقل قرار گرفته است. پارامترهای مستقل ورودی مرتبط با شرایط عملیاتی شامل دمای عملیات، نرخ تغییرات حرارتی، زمان واکنش، و اندازه ذرات، و نرخ سیال حامل ورودی در خصوص متغیرهای وابسته خروجی به عنوان پیش بینی های مدل یادگیری ماشین مواردی همچون عملکرد تولید روغن زیستی، گاز زیستی و کربن زیستی در گروه خروجی های مرتبط با عملکرد کمی فرایند پیرولیز قرار گرفته اند. به منظور پیش بینی کیفیت روغن زیستی تولید شده مواردی همچون بالاترین ارزش حرارتی، نسبت کربن به هیدروژن یا کربن به اکسیژن، میزان سولفور و نیتروژن درون روغن زیستی، اسیدی بودن، مقدار محتوی اسیدی و مقادیر مختلف ترکیب های آلی به عنوان متغیرهای وابسته خروجی در نظر گرفته شده است. یکی از نقاط ضعف این مدل ها بررسی پیرولیز زیست توده در شرایط بدون وجود کاتالیست است. پارامترهای شیمیایی و خصوصیت های فیزیکی و هندسی کاتالیست به مقدار زیادی می تواند فرایند تولید روغن زیستی از زیست توده را تحت تأثیر قرار دهد که مدل سازی آن بسیار با اهمیت است. البته چگونگی ورود داده های کاتالیست و انتخاب عوامل مؤثر اثرگذار هر نوع کاتالیست یکی از مشکل های پیشروی این نوع مدل سازی ها خواهد بود. در جدول های ۱ و ۲ به ترتیب مدل سازی های انجام شده توسط پژوهشگران در خصوص عملکرد روغن زیستی و کیفیت آن آورده شده است. با توجه به این جدول ها، در میان روش های بکار برده شده ماشین یادگیری، شیوه های جمعی در پیش بینی مقدار زیست روغن تولید شده به روش پیرولیز و همچنین پیش بینی کیفیت آنها در تحقیق های بیشتری مورد استفاده قرار گرفته اند. از نظر نحوه عملکرد و دقت. شیوه های شبکه عصبی مصنوعی و شیوه جنگل تصادفی تقریباً در بیشتر موارد بکار برده شده عملکرد و دقت مناسبی از خود نشان داده است. از دیگر مدل های با عملکرد مناسب در این حوزه می توان به ماشین بردار پشتیبان اشاره کرد که در این حوزه از خود عملکرد مناسبی ارائه کرده است.

جدول ۱- برخی پژوهش‌های انجام شده در زمینه استفاده از روش‌های ماشین یادگیری برای پیش‌بینی عملکرد روغن‌زیستی از پیرولیز زیست‌توده.

مرجع	مقدار خطای مدل	مقدار ضریب تعیین	مدل یادگیری ماشین	بیشینه اندازه بانک داده	خروجی‌های مدل	ورودی‌های مدل	نوع زیست‌توده اولیه
(Shen et al., 2022)	$2/54 < \text{MSE} < 16/48$	۰/۸۹ - ۰/۹۲	شبکه عصبی مصنوعی	۱۰۰۲	مقدار عملکرد روغن زیستی مقدار عملکرد گاز زیستی مقدار عملکرد کربن زیستی	مقدار سلولوز، همی سلولوز، لیگنین محتوی کربن، هیدروژن، اکسیژن، نیتروژن و سولفور زیست‌توده اندازه ذرات و بیشینه ارزش حرارتی زیست‌توده دمای عملیات، نرخ افزایش دما، زمان ماند، نرخ جریان نیتروژن	شامل مواد لیگنوسلولوزی، ضایعات کاغذ، فاضلاب انسانی
	-	۰/۹۰ - ۰/۹۳	تقویت گرادیان				
	-	۰/۸۶ - ۰/۹۶	درخت تصمیم				
	-	۰/۸۷ - ۰/۹۳	جنگل تصادفی				
	-	۰/۸۸ - ۰/۹۶	رگرسیون دسته‌بندی				
(Zhang et al., 2022)	$\text{RMSE} < 2/9$	۰/۸۹	جنگل تصادفی	۲۸۲	مقدار عملکرد روغن زیستی	مقدار سلولوز، همی سلولوز، لیگنین محتوی کربن، هیدروژن، اکسیژن، نیتروژن و سولفور زیست‌توده دمای عملیات، نرخ افزایش دما اندازه ذرات	
(Taşar, 2022)	$\text{RMSE} < 7/4$ $\text{MAE} < 5/91$	۰/۱۹	رگرسیون چندخطی	۲۳۸	مقدار عملکرد روغن زیستی	محتوی کربن، هیدروژن، اکسیژن، نیتروژن و سولفور زیست‌توده دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن	گیاهان علفی، لیگنوسلولوزیک و جلبک
	$\text{RMSE} < 3/36$ $\text{MAE} < 2/45$	۰/۸۳	درخت تصمیم				
	$\text{RMSE} < 1/3$ $\text{MAE} < 2/15$	۰/۸۶	مدل رگرسیونی گوسین				
	$\text{RMSE} < 2/1$ $\text{MAE} < 1/55$	۰/۹۳	جنگل تصادفی				
	$\text{RMSE} < 7/26$ $\text{MAE} < 5/99$	۰/۲	رگرسیون چندخطی		مقدار عملکرد روغن زیستی	محتوی مواد فرار، خاکستر و کربن تثبیت شده دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن	
	$\text{RMSE} < 2/81$ $\text{MAE} < 2/14$	۰/۸۸	درخت تصمیم				
	$\text{RMSE} < 2/88$ $\text{MAE} < 1/92$	۰/۸۸	مدل رگرسیونی گوسین				
	$\text{RMSE} < 2/03$ $\text{MAE} < 1/52$	۰/۹۴	جنگل تصادفی				
(Yang et al., 2022a)	$\text{RMSE} < 2/45$	۰/۹۲	جنگل تصادفی	۲۹۲	مقدار روغن تولیدی	محتوی مواد فرار، خاکستر و کربن تثبیت شده دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن مقدار سلولوز، همی سلولوز، لیگنین دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن	لیگنوسلولوزیک
	$\text{RMSE} < 3/25$	۰/۸۵					
	$\text{RMSE} < 2/94$	۰/۸۳					
(Dong et al., 2023)	$4/06 < \text{RMSE} < 4/37$	۰/۷۸ تا ۰/۸۷	جنگل تصادفی	۷۵۸	مقدار عملکرد روغن زیستی	محتوی کربن، هیدروژن، اکسیژن، نیتروژن و سولفور	مجموعه‌ای از زیست‌توده‌های
	$6/2 < \text{RMSE} < 6/9$	۰/۶۳ - ۰/۶۹	الگوریتم تقویتی				



			سازگار شونده		مقدار عملکرد گاز زیستی مقدار عملکرد کربن زیستی	زیست توده محتوی مواد فرار، خاکستر و کربن تثبیت شده دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان سیال حامل	مختلف
	$4/8 < RMSE < 9/4$	$0/74 - 0/82$	گرادبان تقویت شده درخت تصمیم				
	$4/3 < RMSE < 4/6$	$0/79$ تا $0/85$	بیشینه گرادبان تقویت شده بیشینه				
(Hough et al., 2017)		$0/98 - 0/99$	شبکه عصبی مصنوعی		مقدار عملکرد روغن زیستی سبک اولیفاتیک مقدار عملکرد روغن زیستی سنگین اولیفاتیک گاز تولیدی سوخت جامد تولیدی (کربن زیستی)	دمای عملیات، نرخ افزایش دما و نسبت وزنی کربن به هیدروژن در لیگنین	لیگنین
		$0/99$	درخت تصمیم				
(Mathur et al., 2023)	$RMSE=6/27$	$0/26$	رگرسیون چندخطی	۳۱۵	عملکرد روغن زیستی	محتوی کربن، هیدروژن، اکسیژن و نیتروژن زیست توده محتوی مواد فرار، خاکستر و کربن تثبیت شده دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن	زیست توده لیگنوسلولوزیک، ضایعات کشاورزی، پسماند صنایع غذایی،
	$RMSE=3/45$	$0/77$	درخت تصمیم				
	$RMSE=2/99$	$0/79$	جنگل تصادفی				
	$RMSE=2/39$	$0/89$	گرادبان تقویت شده درخت تصمیم				
(Özbay & KÖKTEN, 2019)	$MSE=0/01$	$0/95$	شبکه عصبی مصنوعی	۱۱۹	عملکرد روغن زیستی	محتوی کربن، هیدروژن و اکسیژن زیست توده محتوی مواد فرار، خاکستر و کربن تثبیت شده و محتوی رطوبتی دمای عملیات، زمان عملیات، اندازه ذرات و نوع کاتالیست و مقدار کاتالیست	زیست توده چوبی
(Madhu et al., 2017)	$0/28 < RMSE < 0/69$	$0/99$	شبکه عصبی مصنوعی		مقدار عملکرد روغن زیستی مقدار عملکرد گاز زیستی مقدار عملکرد کربن زیستی	دمای عملیات، زمان عملیات، اندازه ذرات، نرخ جریان گاز حامل	ضایعات کتان
(Chen et al., 2018)	$3/35 < RMSE < 4/5$	$0/80 - 0/89$	شبکه عصبی مصنوعی	۱۷۵	مقدار عملکرد روغن زیستی مقدار عملکرد گاز زیستی مقدار عملکرد کربن زیستی	محتوی کربن، هیدروژن، اکسیژن و نیتروژن زیست توده محتوی خاکستر و رطوبت زیست توده دمای عملیات، نرخ جریان سیال و اندازه ذرات	انواع مختلف زیست توده
	$2/53 < RMSE < 3/34$	$0/89$	ماشین بردار پشتیبان				
(Shahbeik et al., 2022)	$0/21 < RMSE < 12/51$	$0/65 - 0/94$	جنگل تصادفی	۲۴۴	مقدار عملکرد روغن - زیستی مقدار عملکرد گاز زیستی مقدار عملکرد کربن - زیستی	محتوی کربن، هیدروژن، اکسیژن و نیتروژن زیست توده محتوی مواد فرار، خاکستر و کربن تثبیت شده دمای عملیات، زمان عملیات و نرخ افزایش دما	زیست توده فاضلاب انسانی و حیوانی

جدول ۲- برخی پژوهش‌های انجام شده در زمینه استفاده از روش‌های ماشین یادگیری برای پیش‌بینی کیفیت روغن زیستی به‌دست‌آمده از فرایند پیرولیز زیست‌توده.

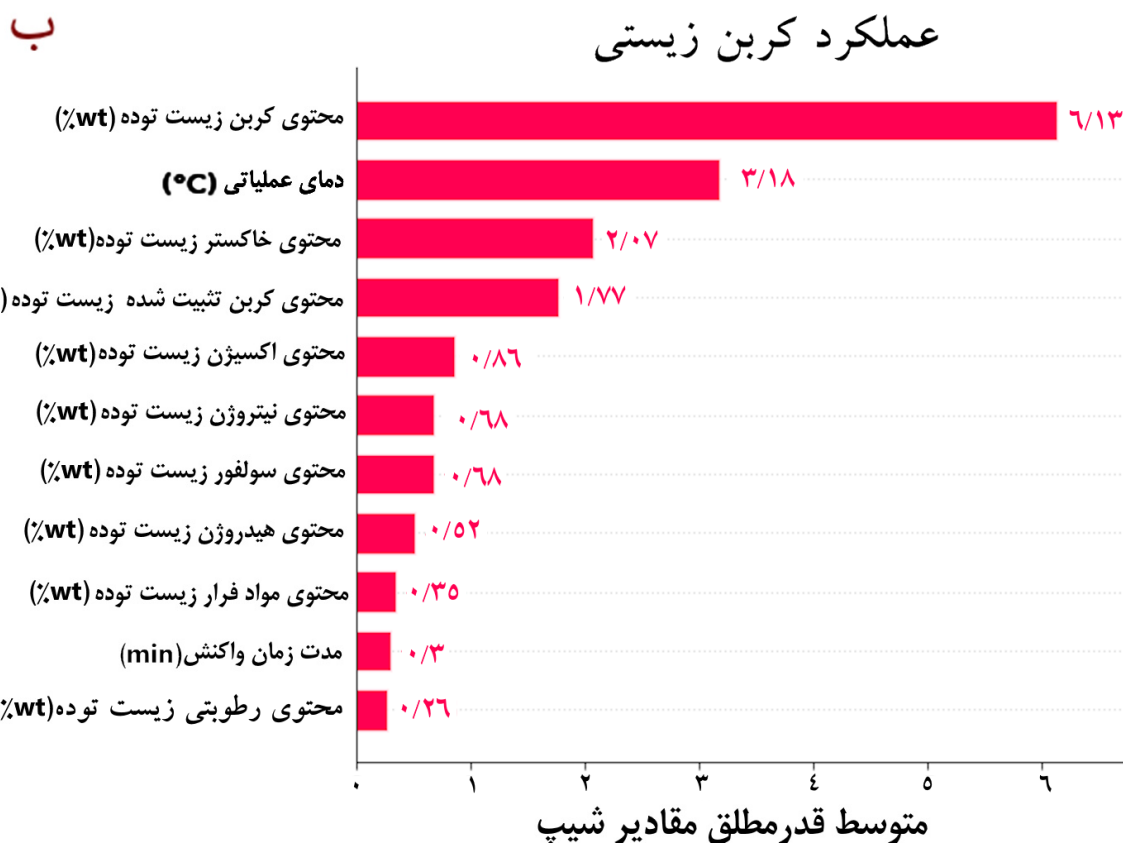
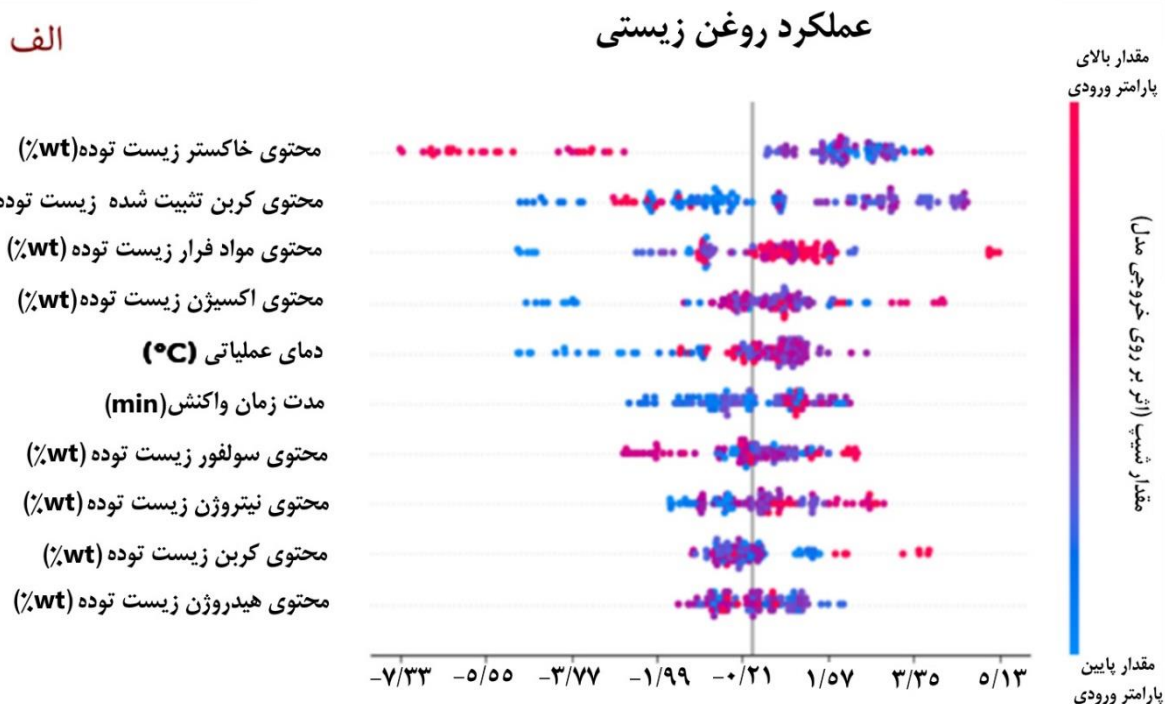
مرجع	مقدار خطای مدل	مقدار ضریب تعیین	مدل یادگیری ماشین	اندازه بانک داده	خروجی‌های مدل	ورودی‌های مدل	نوع زیست‌توده اولیه
(Zhang et al., 2022)	RMSE < ۰/۰۱۵	۰/۷۲	جنگل تصادفی	۲۳۲	روان روی روغن زیستی	محتوی کربن، هیدروژن، اکسیژن، نیتروژن و سولفور زیست‌توده	لیگنوسلولوزیک
	RMSE < ۰/۱۴	۰/۷۸			نسبت اکسیژن به کربن روغن زیستی بالاترین ارزش حرارتی		
	RMSE < ۱/۶۹	۰/۶۷			نسبت هیدروژن به کربن روغن زیستی		
	RMSE < ۰/۶۴	۰/۷۵					
(Taşar, 2022)	RMSE < ۰/۹۱ MAE < ۰/۷۴	۰/۶	رگرسیون چندخطی	۶۰	مقدار هیدروژن روغن زیستی	محتوی کربن، هیدروژن، اکسیژن، نیتروژن و سولفور زیست‌توده دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن	گیاهان علفی، لیگنوسلولوزیک و چلبک
	RMSE < ۰/۸۶ MAE < ۰/۶۳	۰/۶۵	درخت تصمیم				
	RMSE < ۰/۹۶ MAE < ۰/۷۳	۰/۵۶	مدل رگرسیونی گوسین				
	RMSE < ۰/۶۲ MAE < ۰/۴۴	۰/۸۲	جنگل تصادفی				
	RMSE < ۰/۹۱ MAE < ۰/۷۳	۰/۶	رگرسیون چندخطی		مقدار هیدروژن روغن زیستی	محتوی مواد فرار، خاکستر و کربن تثبیت شده دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن	
	RMSE < ۰/۸۳ MAE < ۰/۶۴	۰/۶۷	درخت تصمیم				
	RMSE < ۰/۸۲ MAE < ۰/۶۶	۰/۶۸	مدل رگرسیونی گوسین				
	RMSE < ۰/۴۶ MAE < ۰/۳۵	۰/۹	جنگل تصادفی				
(Yang et al., 2022a)	RMSE < ۲/۵۴	۰/۷۹	جنگل تصادفی	۱۷۱	مقدار اکسیژن روغن زیستی	محتوی مواد فرار، خاکستر و کربن تثبیت شده دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن	لیگنوسلولوزیک
	RMSE < ۲/۹۸	۰/۸۲				مقدار سلولوز، لیگنین دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن	
	RMSE < ۲/۲۷	۰/۸۹				محتوی مواد فرار، خاکستر و کربن تثبیت شده دمای عملیات، نرخ افزایش دما و اندازه ذرات و نرخ جریان نیتروژن	
(Su & Wang, 2023)	RMSE < ۴/۰۱	۰/۹۱	بیشینه‌گرادیان تقویت شده بیشینه	۱۲۲	محتوی اسیدی روغن زیستی	محتوی کربن، هیدروژن، اکسیژن و نیتروژن زیست‌توده دمای عملیات، نرخ جریان سیال و اندازه ذرات	بقایای جنگلی، باقی‌مانده غذا، ضایعات کشاورزی و همچنین زیست‌توده‌های پیش‌فراوری شده از روش‌های
	RMSE < ۶/۷۶	۰/۸۲			محتوی فنلی روغن زیستی		
	RMSE < ۴/۲۳	۰/۸۵			محتوی کتونی و آلدئیدی روغن زیستی		

	RMSE < ۳/۹۲	۰/۸۲			محتوی اسیدی روغن زیستی	محتوی مواد فرار، خاکستر و کربن تثبیت شده دمای عملیات، نرخ جریان سیال و اندازه ذرات	اسیدشویی شده، پخته شده
	RMSE < ۷/۰۹	۰/۷۲			محتوی فنلی روغن زیستی		
	RMSE < ۴/۹۹	۰/۸۵			محتوی کتونی و آلدئیدی روغن زیستی		
(Chen et al., 2018)	RMSE < ۱/۰۳ MAE < ۰/۷۵	۰/۹۳	شبکه عصبی مصنوعی	۷۱	بالاترین ارزش حرارتی روغن زیستی	محتوی کربن، هیدروژن، اکسیژن و نیتروژن زیست توده محتوی خاکستر و رطوبت زیست توده دمای عملیات، نرخ جریان سیال و اندازه ذرات	انواع مختلف زیست توده
	RMSE < ۰/۹۸ MAE < ۰/۷۶	۰/۹۳	ماشین بردار پشتیبان				

کاربرد تحلیل اهمیت پارامتر در یادگیری ماشین برای پیش بینی فرایند تولید و کیفیت روغن زیستی

با وجود اینکه تکنیک‌های یادگیری ماشین قادرند پدیده‌های پیچیده مانند پیرولیز را مدل‌سازی کنند، اما طبیعت تکنیک‌های یادگیری ماشین که به صورت یک جعبه سیاه است باعث می‌شود تفسیر نتایج و عوامل اثرگذار بر روی مدل بسیار چالش برانگیز باشد. لذا، نتایج هر مدل یادگیری ماشین باید با استفاده از ابزارهای پیشرفته توضیح داده شوند. یک رویکرد محبوب برای تفسیر مدل‌های یادگیری ماشین و تحلیل اهمیت پارامتر روش شیپ^۱ است. این شیوه راهی برای اندازه‌گیری تأثیر هر پارامتر مستقل ورودی بر پاسخ وابسته در خروجی مدل یادگیری ماشین فراهم می‌کند. به این ترتیب میزان اهمیت، میانگین مقدار مشارکت در شکل دادن مدل، جهت اثرگذاری (مثبت یا منفی) با استفاده از این شیوه به دست می‌آید (Shafizadeh et al., 2023c). این روش براساس نظریه بازی‌ها ارتقا یافته است تا به کشف روابط پیچیده در معادله‌های غیرخطی مدل‌سازی‌های یادگیری ماشین بپردازد. این روش اثر هر پارامتر را به شیوه‌های مختلف مانند متوسط ضریب اثرگذاری یا بیشینه مقدار اثرگذاری می‌تواند ارائه کند. از نمودارهای مرسوم که برای تحلیل در این شیوه استفاده می‌شود می‌توان به نمودارهای طرح زنبوری و نمودار قدرمطلق بیشینه اشاره کرد (Shafizadeh et al., 2023c).

اخیرا چندین مطالعه از تحلیل حساسیت شیپ برای بررسی پارامترهای موثر در مدل‌سازی فرایند حرارتی-شیمیایی استفاده شده است. این شیوه برای تحلیل حساسیت مدل‌سازی (روش گرادیان تقویت شده درخت تصمیم) انجام شده در پیرولیز زیست توده با انرژی امواج مایکروویو بکار گرفته شده است (Yang et al., 2022b). از این روش در مدل‌سازی پیرولیز زیست توده (فاضلاب) به روش یادگیری ماشین استفاده شده است که نتایج تحلیل حساسیت (شکل ۶) تطابق مناسبی با آزمایش‌های تجربی صورت گرفته نشان داده است (Shahbeik et al., 2022). شکل ۶-الف نمودار طرح زنبوری تحلیل اهمیت را نشان می‌دهد. در این نمودار هر میله افقی نشان‌گر یک پارامتر ورودی مدل‌سازی است. رنگ نقاط بر روی هر میله بیانگر مقدار بیشینه یا کمینه وارد شده آن پارامتر در مدل‌سازی است (مقادیر قرمز بیشینه و مقادیر آبی کمینه) همچنین محل قرارگیری این نقاط بر روی محور افقی نشان‌دهنده میزان و جهت اثرگذاری بر روی خروجی مدل است به نحوی که مقادیر قرارگرفته در سمت راست محور با مبداء صفر اثر مثبت و مقادیر قرار گرفته در سمت چپ آن اثر منفی پارامتر بر روی مدل را نشان می‌دهند. به عنوان نمونه مقدار خاکستر زیست توده در رده مهمترین پارامترهای اثرگذار بر روی تولید روغن زیستی در طول فرایند مدل‌سازی پیرولیز زیست توده فاضلاب قرار گرفته است. نمودار طرح زنبوری این پارامتر نشان داده است که مقادیر زیاد محتوی خاکستر اثر منفی بر روی عملکرد روغن زیستی تولیدی در طول فرایند مدل‌سازی داشته است. علاوه بر نمودارهای طرح زنبوری نمودارهای قدرمطلق متوسط مقدار شیپ نیز برای یافتن پارامترها با بیشترین مقدار متوسط تأثیر بر روی خروجی یک مدل‌سازی مد نظر استفاده می‌شوند. در این نمودارها که در شکل ۶-ب نشان داده شده است امکان دریافت تأثیر مثبت یا منفی بر روی یک خروجی مشخص وجود ندارد اما میزان متوسط قدرمطلق تأثیر بر روی آن خروجی مد نظر قابل شناسایی است.



شکل ۶. یک نمونه از تحلیل اهمیت پارامتر (شیپ) در مدل تولید شده برای تبدیل حرارتی-شیمیایی زیست توده به شیوه ماشین یادگیری. الف) نمودار طرح زنبوری شیپ (Shahbeik et al., 2022)، ب) نمودار قدر مطلق متوسط مقدار شیپ (Shafizadeh et al., 2023c).

نتیجه گیری و راهبردهای آینده

باتوجه به نیاز روزافزون بشر به انرژی و همزمان نیاز مبرم به کاهش اثرات ناشی از گرمایش زمین استفاده از انرژی‌های تجدیدپذیر، گریزناپذیر است. استفاده از زیست توده به عنوان یک منبع تجدیدپذیر و پاک در این زمینه مورد توجه گسترده قرار گرفته است. روش‌های

مختلفی برای تبدیل زیست‌توده به انرژی به کار گرفته شده است. در این میان روش‌های حرارتی - شیمیایی و بخصوص پیرولیز از اقبال خاصی برای تولید روغن زیستی برخوردار بوده‌اند. روش‌های شیمیایی حرارتی از پیچیدگی بسیار زیادی برخوردار هستند و عوامل متعددی چون ترکیب زیست‌توده و شرایط عملیاتی بر روی آنها اثرگذار است. به‌دست‌آوردن شرایط بهینه برای تولید بهینه روغن زیستی از نظر عملکرد و کیفیت یکی از عوامل نیازهای محققان برای توسعه این زمینه است. با توجه به زمان‌بر بودن و هزینه‌موردنیاز زیاد برای انجام آزمایش‌های تجربی جهت یافتن عوامل مهم در پیرولیز و چگونگی اثرگذاری آنها بر روی فرایند استفاده از روش‌های مدل‌سازی می‌تواند بسیار اثربخش و کمک‌کننده باشد. در میان روش‌های مدل‌سازی، شیوه‌های مبتنی بر یادگیری ماشین در سال‌های اخیر مورد توجه بسیاری قرار گرفته‌اند. این امر به خاطر نیاز محاسباتی به مراتب کمتر و پیش‌نیازهای مناسب‌تر (وجود داده‌های ورودی مستقل و خروجی وابسته) و عدم نیاز به ساده‌سازی و فرض‌های متعدد پیش از مدل‌سازی نسبت به شیوه‌های دیگر است. بر این اساس تحقیق‌های زیادی بخصوص از سال ۲۰۱۹ در زمینه کاربرد ماشین یادگیری در پیرولیز زیست‌توده بکار رفته است. در میان روش‌های مختلف ماشین یادگیری از روش‌های ماشین‌یادگیری جمعی مانند جنگل تصادفی در تحقیق‌های بیشتری برای پیش‌بینی عملکرد روغن زیستی و کیفیت آن بکار رفته‌اند و از نظر دقت نیز روش جنگل تصادفی و شبکه عصبی مصنوعی عملکرد بهتری از نظر دقت و کم‌بودن خطاهای پیش‌بینی از خود نشان داده‌اند. علاوه بر این موارد امکان ارائه تحلیل‌های حساسیت بر روی تک‌تک پارامترهای مستقل ورودی و نحوه اثرگذاری آنها بر روی عملکرد کلی مدل ارتقا یافته باعث مقبولیت بیشتر این استفاده از ماشین یادگیری در این زمینه شده است. بدین ترتیب امکان تفسیر مدل‌های ارتقا یافته و مقایسه نتایج خروجی با شرایط واقعی و تفسیر آنها امکان‌پذیر می‌گردد.

با توجه به قابلیت‌های گسترده ماشین یادگیری در مدل‌سازی فرایندهای پیچیده امکان توسعه و گسترش این مدل‌ها به فرایندهای پیچیده‌تر وجود دارد. از نقطه ضعف مدل‌های ارائه شده محدود بودن بانک داده مورد استفاده مناسب برای ارائه یک مدل کلی برای فرایند پیرولیز زیست‌توده است؛ بنابراین توصیه می‌گردد ارتقای بانک داده که در برگیرنده طیف گسترده‌ای از زیست‌توده‌های مختلف باشد در آینده مدنظر قرار بگیرد. علاوه بر این می‌توان با ارتقای بانک داده برای انواع زیست‌توده مانند انواع لیگنوسلولزیک و یا غیر لیگنوسلولزیک مدل‌های مختص به هر کدام از دسته‌ها را ارتقا داد. از دیگر نقطه‌ضعف‌های مدل‌های ارائه شده تاکنون عدم ورود بسیاری از پارامترهای مؤثر بر روی فرایند پیرولیز مانند استفاده از کاتالیست‌ها است. این مسئله در تحقیق‌های آینده بایستی مدنظر قرار بگیرد. روش‌های بکار برده شده تاکنون به مدل‌سازی فرایند پیرولیز زیست‌توده در شرایط حضور کاتالیست نپرداخته‌اند. استفاده از کاتالیست، نوع آن، مقدار بکار برده شده و خصوصیات ساختاری و شیمیایی آن بر روی عملکرد تولید روغن زیستی و بهبود کیفیت آن بسیار اثرگذار است؛ بنابراین ارائه راه‌حلی در آینده که بتواند به صورت کمی ویژگی‌های مهم کاتالیست را در طول فرایند مدل‌سازی به‌عنوان پارامتر مستقل وارد نماید و اثر این عوامل بر روی خروجی‌ها را به نحو مطلوبی مدل کند بسیار با اهمیت خواهد بود. محدودیت دیگر پیشروی مدل‌سازی بررسی امکان مدل‌سازی اثر کاتالیست‌های پیچیده که از چند ماده مختلف مانند پایه کاتالیست، شتاب‌دهنده کاتالیست^۱، در ترکیب با کاتالیست‌های فلزی یا غیرفلزی است. چگونگی دخیل نمودن این موارد در مدل‌سازی یکی از چالش‌های آینده در این زمینه خواهد بود. علاوه بر این مدل‌سازی‌های آینده می‌تواند کاربرد کاتالیست به صورت درون‌جا و برون‌جا را به طور مجزا بررسی نماید. استفاده از روش‌های به‌روز ارائه شده برای مدل‌های ماشین یادگیری که امکان مدل نمودن مسائل پیچیده‌تر دارد، برای مدل‌کردن فرایندهای پیچیده پیرولیز کاتالیستی بایستی برای تحقیق‌های آینده مدنظر قرار گیرد. همچنین تاکنون در مدل‌سازی‌های ارائه شده و بررسی شده از روش‌های بهینه‌سازی چندهدفه^۲ برای یافتن نقاط بهینه عملیاتی و یا ویژگی‌های بهینه کاتالیست و زیست‌توده جهت دستیابی به یک روغن زیستی با ترکیب خاص و کیفیت مدنظر استفاده نشده است؛ بنابراین استفاده از این روش‌های بهینه‌سازی می‌تواند به محقق کمک کند تا محدوده نقاط بهینه عملیاتی برای دستیابی به روغن زیستی با ویژگی‌های خاص (مثلاً مناسب تولید انرژی یا جهت تولید محصول‌های شیمیایی با ارزش بالا) را پیش از شروع انجام آزمایش‌های تجربی، براساس مدل‌سازی‌های انجام شده در دست داشته باشد. از دیگر کاربردهای مدل‌سازی‌ها ماشین یادگیری در آینده می‌توان به کنترل برخط راکتورهای آزمایشگاهی پیرولیز زیست‌توده جهت انجام آزمایش‌های تکمیلی برای بهبود فرایند تولید روغن زیستی اشاره نمود. علاوه بر این استفاده از مدل‌سازی به‌وسیله ماشین یادگیری در سامانه‌های نیمه‌صنعتی و صنعتی شده پیرولیز زیست‌توده به‌منظور بهبود عملکرد این واحدها می‌تواند مورد بررسی قرار گیرد.

1 Support

2 Promoter

3 Multi objective optimization

REFERENCES

- Abnisa, F., Arami-Niya, A., Wan Daud, W. M. A., Sahu, J. N., & Noor, I. M. (2013). Utilization of oil palm tree residues to produce bio-oil and bio-char via pyrolysis. *Energy Conversion and Management*, 76, 1073–1082. <https://doi.org/10.1016/j.enconman.2013.08.038>
- Akhtar, J., & Saidina Amin, N. (2012). A review on operating parameters for optimum liquid oil yield in biomass pyrolysis. *Renewable and Sustainable Energy Reviews*, 16(7), 5101–5109. <https://doi.org/10.1016/j.rser.2012.05.033>
- Al-Sabawi, M., Chen, J., & Ng, S. (2012). Fluid Catalytic Cracking of Biomass-Derived Oils and Their Blends with Petroleum Feedstocks: A Review. *Energy & Fuels*, 26(9), 5355–5372. <https://doi.org/10.1021/ef3006417>
- Awasthi, M. K., Sindhu, R., Sirohi, R., Kumar, V., Ahluwalia, V., Binod, P., Juneja, A., Kumar, D., Yan, B., Sarsaiya, S., Zhang, Z., Pandey, A., & Taherzadeh, M. J. (2022). Agricultural waste biorefinery development towards circular bioeconomy. *Renewable and Sustainable Energy Reviews*, 158, 112122. <https://doi.org/10.1016/j.rser.2022.112122>
- Basu, P. (2018). *Biomass Gasification, Pyrolysis and Torrefaction*. Elsevier. <https://doi.org/10.1016/C2016-0-04056-1>
- Campbell, R. M., Anderson, N. M., Daugaard, D. E., & Naughton, H. T. (2018). Financial viability of biofuel and biochar production from forest biomass in the face of market price volatility and uncertainty. *Applied Energy*, 230, 330–343. <https://doi.org/10.1016/j.apenergy.2018.08.085>
- Carpenter, D., Westover, T. L., Czernik, S., & Jablonski, W. (2014). Biomass feedstocks for renewable fuel production: a review of the impacts of feedstock and pretreatment on the yield and product distribution of fast pyrolysis bio-oils and vapors. *Green Chem.*, 16(2), 384–406. <https://doi.org/10.1039/C3GC41631C>
- Chakraborty, B., & Murphy, S. A. (2014). Dynamic Treatment Regimes. *Annual Review of Statistics and Its Application*, 1(1), 447–464. <https://doi.org/10.1146/annurev-statistics-022513-115553>
- Chen, D., Cen, K., Zhuang, X., Gan, Z., Zhou, J., Zhang, Y., & Zhang, H. (2022). Insight into biomass pyrolysis mechanism based on cellulose, hemicellulose, and lignin: Evolution of volatiles and kinetics, elucidation of reaction pathways, and characterization of gas, biochar and bio-oil. *Combustion and Flame*, 242, 112142. <https://doi.org/10.1016/j.combustflame.2022.112142>
- Chen, W.-H., Lin, B.-J., Lin, Y.-Y., Chu, Y.-S., Ubando, A. T., Show, P. L., Ong, H. C., Chang, J.-S., Ho, S.-H., Culaba, A. B., Pétrissans, A., & Pétrissans, M. (2021). Progress in biomass torrefaction: Principles, applications and challenges. *Progress in Energy and Combustion Science*, 82, 100887. <https://doi.org/10.1016/j.pecs.2020.100887>
- Chen, X., Zhang, H., Song, Y., & Xiao, R. (2018). Prediction of product distribution and bio-oil heating value of biomass fast pyrolysis. *Chemical Engineering and Processing - Process Intensification*, 130(February), 36–42. <https://doi.org/10.1016/j.cep.2018.05.018>
- Crisci, C., Ghattas, B., & Perera, G. (2012). A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling*, 240, 113–122. <https://doi.org/10.1016/j.ecolmodel.2012.03.001>
- De Clercq, R., Dusselier, M., & Sels, B. F. (2017). Heterogeneous catalysis for bio-based polyester monomers from cellulosic biomass: advances, challenges and prospects. *Green Chem.*, 19(21), 5012–5040. <https://doi.org/10.1039/C7GC02040F>
- Di Blasi, C., Galgano, A., & Branca, C. (2009). Influences of the Chemical State of Alkaline Compounds and the Nature of Alkali Metal on Wood Pyrolysis. *Industrial & Engineering Chemistry Research*, 48(7), 3359–3369. <https://doi.org/10.1021/ie801468y>
- Ding, Y., Ezekoye, O. A., Zhang, J., Wang, C., & Lu, S. (2018). The effect of chemical reaction kinetic parameters on the bench-scale pyrolysis of lignocellulosic biomass. *Fuel*, 232, 147–153. <https://doi.org/https://doi.org/10.1016/j.fuel.2018.05.140>
- Dong, Z., Bai, X., Xu, D., & Li, W. (2023). Machine learning prediction of pyrolytic products of lignocellulosic biomass based on physicochemical characteristics and pyrolysis conditions. *Bioresource Technology*, 367, 128182. <https://doi.org/10.1016/j.biortech.2022.128182>
- Foong, S. Y., Liew, R. K., Yang, Y., Cheng, Y. W., Yek, P. N. Y., Wan Mahari, W. A., Lee, X. Y., Han, C. S., Vo, D.-V. N., Van Le, Q., Aghbashlo, M., Tabatabaei, M., Sonne, C., Peng, W., & Lam, S. S. (2020). Valorization of biomass waste to engineered activated biochar by microwave pyrolysis: Progress, challenges, and future directions. *Chemical Engineering Journal*, 389, 124401. <https://doi.org/10.1016/j.cej.2020.124401>



- Gani, A., & Naruse, I. (2007). Effect of cellulose and lignin content on pyrolysis and combustion characteristics for several types of biomass. *Renewable Energy*, 32(4), 649–661. <https://doi.org/10.1016/j.renene.2006.02.017>
- Gautam, P., Neha, Upadhyay, S. N., & Dubey, S. K. (2020). Bio-methanol as a renewable fuel from waste biomass: Current trends and future perspective. *Fuel*, 273, 117783. <https://doi.org/10.1016/j.fuel.2020.117783>
- Ghafarian, F., Wieland, R., Lüttschwager, D., & Nendel, C. (2022). Application of extreme gradient boosting and Shapley Additive explanations to predict temperature regimes inside forests from standard open-field meteorological data. *Environmental Modelling & Software*, 156, 105466. <https://doi.org/10.1016/j.envsoft.2022.105466>
- Harrison, J. H., Gilbertson, J. R., Hanna, M. G., Olson, N. H., Seheult, J. N., Sorace, J. M., & Stram, M. N. (2021). Introduction to Artificial Intelligence and Machine Learning for Pathology. *Archives of Pathology & Laboratory Medicine*, 145(10), 1228–1254. <https://doi.org/10.5858/arpa.2020-0541-CP>
- Hassan, E. M., Steele, P. H., & Ingram, L. (2009). Characterization of Fast Pyrolysis Bio-oils Produced from Pretreated Pine Wood. *Applied Biochemistry and Biotechnology*, 154(1–3), 3–13. <https://doi.org/10.1007/s12010-008-8445-3>
- Haykiri-Acma, H. (2006). The role of particle size in the non-isothermal pyrolysis of hazelnut shell. *Journal of Analytical and Applied Pyrolysis*, 75(2), 211–216. <https://doi.org/10.1016/j.jaap.2005.06.002>
- Hodge, V. J., & Austin, J. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22(2), 85–126. <https://doi.org/10.1007/s10462-004-4304-y>
- Hough, B. R., Beck, D. A. C., Schwartz, D. T., & Pfaendtner, J. (2017). Application of machine learning to pyrolysis reaction networks: Reducing model solution time to enable process optimization. *Computers and Chemical Engineering*, 104, 56–63. <https://doi.org/10.1016/j.compchemeng.2017.04.012>
- Kaczor, Z., Buliński, Z., & Werle, S. (2020). Modelling approaches to waste biomass pyrolysis: a review. *Renewable Energy*, 159, 427–443. <https://doi.org/10.1016/j.renene.2020.05.110>
- Kasmuri, N. H., Kamarudin, S. K., Abdullah, S. R. S., Hasan, H. A., & Som, A. M. (2019). Integrated advanced nonlinear neural network-simulink control system for production of bio-methanol from sugar cane bagasse via pyrolysis. *Energy*, 168(2019), 261–272. <https://doi.org/10.1016/j.energy.2018.11.056>
- Khan, M., Raza Naqvi, S., Ullah, Z., Ali Ammar Taqvi, S., Nouman Aslam Khan, M., Farooq, W., Taqi Mehran, M., Juchelková, D., & Štěpanec, L. (2023). Applications of machine learning in thermochemical conversion of biomass-A review. *Fuel*, 332, 126055. <https://doi.org/10.1016/j.fuel.2022.126055>
- Kim, D., & Philen, M. (2011). *Damage classification using Adaboost machine learning for structural health monitoring* (M. Tomizuka (ed.); p. 79812A). <https://doi.org/10.1117/12.882016>
- Kumar, P., Barrett, D. M., Delwiche, M. J., & Stroeve, P. (2009). Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. *Industrial & Engineering Chemistry Research*, 48(8), 3713–3729. <https://doi.org/10.1021/ie801542g>
- Kumar Sharma, A., Kumar Ghodke, P., Goyal, N., Nethaji, S., & Chen, W.-H. (2022). Machine learning technology in biohydrogen production from agriculture waste: Recent advances and future perspectives. *Bioresource Technology*, 364, 128076. <https://doi.org/10.1016/j.biortech.2022.128076>
- Leng, E., He, B., Chen, J., Liao, G., Ma, Y., Zhang, F., Liu, S., & E, J. (2021). Prediction of three-phase product distribution and bio-oil heating value of biomass fast pyrolysis based on machine learning. *Energy*, 236, 121401. <https://doi.org/10.1016/j.energy.2021.121401>
- Li, S., Cheng, S., & Cross, J. S. (2020). Homogeneous and Heterogeneous Catalysis Impact on Pyrolyzed Cellulose to Produce Bio-Oil. *Catalysts*, 10(2), 178. <https://doi.org/10.3390/catal10020178>
- Liakos, K., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors*, 18(8), 2674. <https://doi.org/10.3390/s18082674>
- Liu, C., Wang, H., Karim, A. M., Sun, J., & Wang, Y. (2014). Catalytic fast pyrolysis of lignocellulosic biomass. *Chem. Soc. Rev.*, 43(22), 7594–7623. <https://doi.org/10.1039/C3CS60414D>
- Liu, M., Liu, X., Li, J., Ding, C., & Jiang, J. (2014). Evaluating total inorganic nitrogen in coastal waters through fusion of multi-temporal RADARSAT-2 and optical imagery using random forest algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 33, 192–202. <https://doi.org/10.1016/j.jag.2014.05.009>
- Madadian, E., Haelssig, J. B., & Pegg, M. (2020). A Comparison of Thermal Processing Strategies for Landfill Reclamation: Methods, Products, and a Promising Path Forward. *Resources, Conservation and Recycling*, 160, 104876. <https://doi.org/10.1016/j.resconrec.2020.104876>
- Madhu, P., Matheswaran, M. M., & Periyanyagi, G. (2017). Optimization and characterization of bio-oil

- produced from cotton shell by flash pyrolysis using artificial neural network. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 39(23), 2173–2180. <https://doi.org/10.1080/15567036.2017.1403508>
- Mateus, M. M., Bordado, J. M., & Galhano dos Santos, R. (2021). Estimation of higher heating value (HHV) of bio-oils from thermochemical liquefaction by linear correlation. *Fuel*, 302, 121149. <https://doi.org/10.1016/j.fuel.2021.121149>
- Mathur, J., Baruah, B., & Tiwari, P. (2023). Prediction of bio-oil yield during pyrolysis of lignocellulosic biomass using machine learning algorithms. *The Canadian Journal of Chemical Engineering*, 101(5), 2457–2471. <https://doi.org/10.1002/cjce.24674>
- Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1(4), 140–147. <https://doi.org/10.38094/jastt1457>
- Murali, N., Kucukkaya, A., Petukhova, A., Onofrey, J., & Chapiro, J. (2020). Supervised Machine Learning in Oncology: A Clinician's Guide. *Digestive Disease Interventions*, 04(01), 073–081. <https://doi.org/10.1055/s-0040-1705097>
- Nie, P., Roccotelli, M., Fanti, M. P., Ming, Z., & Li, Z. (2021). Prediction of home energy consumption based on gradient boosting regression tree. *Energy Reports*, 7, 1246–1255. <https://doi.org/10.1016/j.egyr.2021.02.006>
- Norouzi, O., Taghavi, S., Arku, P., Jafarian, S., Signoretto, M., & Dutta, A. (2021). What is the best catalyst for biomass pyrolysis? *Journal of Analytical and Applied Pyrolysis*, 158, 105280. <https://doi.org/10.1016/j.jaap.2021.105280>
- Özbay, G., & KÖKTEN, E. S. (2019). Modeling of Bio-Oil Production by Pyrolysis of Woody Biomass: Artificial Neural Network Approach. *Journal of Polytechnic*, 0900, 0–3. <https://doi.org/10.2339/politeknik.659136>
- Pütün, A. E., Özbay, N., Apaydin Varol, E., Uzun, B. B., & Ateş, F. (2007). Rapid and slow pyrolysis of pistachio shell: effect of pyrolysis conditions on the product yields and characterization of the liquid product. *International Journal of Energy Research*, 31(5), 506–514. <https://doi.org/10.1002/er.1263>
- Raghavendra, N. S., & Deka, P. C. (2014). Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, 19, 372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>
- Rahman, M. M., Liu, R., & Cai, J. (2018). Catalytic fast pyrolysis of biomass over zeolites for high quality bio-oil – A review. *Fuel Processing Technology*, 180, 32–46. <https://doi.org/10.1016/j.fuproc.2018.08.002>
- Rashidi, H. H., Albahra, S., Robertson, S., Tran, N. K., & Hu, B. (2023). Common statistical concepts in the supervised Machine Learning arena. *Frontiers in Oncology*, 13. <https://doi.org/10.3389/fonc.2023.1130229>
- Rzychoń, M., Żogała, A., & Róg, L. (2021). Experimental study and extreme gradient boosting (XGBoost) based prediction of caking ability of coal blends. *Journal of Analytical and Applied Pyrolysis*, 156, 105020. <https://doi.org/10.1016/j.jaap.2021.105020>
- Salehi, E., Abedi, J., & Harding, T. (2009). Bio-oil from Sawdust: Pyrolysis of Sawdust in a Fixed-Bed System. *Energy & Fuels*, 23(7), 3767–3772. <https://doi.org/10.1021/ef900112b>
- Samanpour, A. R., Ruegenberg, A., & Ahlers, R. (2018). The Future of Machine Learning and Predictive Analytics. In *Digital Marketplaces Unleashed* (pp. 297–309). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49275-8_30
- Seo, M. W., Lee, S. H., Nam, H., Lee, D., Tokmurzin, D., Wang, S., & Park, Y.-K. (2022). Recent advances of thermochemical conversion processes for biorefinery. *Bioresource Technology*, 343, 126109. <https://doi.org/10.1016/j.biortech.2021.126109>
- Shafizadeh, A., Rastegari, H., Shahbeik, H., Mobli, H., Pan, J., Peng, W., Li, G., Tabatabaei, M., & Aghbashlo, M. (2023a). A critical review of the use of nanomaterials in the biomass pyrolysis process. *Journal of Cleaner Production*, 400, 136705. <https://doi.org/10.1016/j.jclepro.2023.136705>
- Shafizadeh, A., Shahbeig, H., Hossein Nadian, M., Mobli, H., Dowlati, M., Kumar Gupta, V., Peng, W., Shiung Lam, S., Tabatabaei, M., & Aghbashlo, M. (2022). Machine learning predicts and optimizes hydrothermal liquefaction of biomass. *Chemical Engineering Journal*, 136579. <https://doi.org/10.1016/j.cej.2022.136579>
- Shafizadeh, A., Shahbeik, H., Nadian, M. H., Gupta, V. K., Nizami, A.-S., Lam, S. S., Peng, W., Pan, J., Tabatabaei, M., & Aghbashlo, M. (2023b). Turning hazardous volatile matter compounds into fuel by catalytic steam reforming: An evolutionary machine learning approach. *Journal of Cleaner Production*,



- 413, 137329. <https://doi.org/10.1016/j.jclepro.2023.137329>
- Shafizadeh, A., Shahbeik, H., Rafiee, S., Moradi, A., Shahbaz, M., Madadi, M., Li, C., Peng, W., Tabatabaei, M., & Aghbashlo, M. (2023c). Machine learning-based characterization of hydrochar from biomass: Implications for sustainable energy and material production. *Fuel*, 347, 128467. <https://doi.org/10.1016/j.fuel.2023.128467>
- Shahbeik, H., Rafiee, S., Shafizadeh, A., Jeddi, D., Jafary, T., Lam, S. S., Pan, J., Tabatabaei, M., & Aghbashlo, M. (2022). Characterizing sludge pyrolysis by machine learning: Towards sustainable bioenergy production from wastes. *Renewable Energy*, 199, 1078–1092. <https://doi.org/10.1016/j.renene.2022.09.022>
- Shahbeik, H., Shafizadeh, A., Gupta, V. K., Lam, S. S., Rastegari, H., Peng, W., Pan, J., Tabatabaei, M., & Aghbashlo, M. (2023). Using nanocatalysts to upgrade pyrolysis bio-oil: A critical review. *Journal of Cleaner Production*, 413, 137473. <https://doi.org/10.1016/j.jclepro.2023.137473>
- Shen, J., Yan, M., Fang, M., & Gao, X. (2022). Machine learning-based modeling approaches for estimating pyrolysis products of varied biomass and operating conditions. *Bioresource Technology Reports*, 20, 101285. <https://doi.org/10.1016/j.biteb.2022.101285>
- Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016). A review of machine learning techniques using decision tree and support vector machine. *2016 International Conference on Computing Communication Control and Automation (ICCUBEA)*, 1–7. <https://doi.org/10.1109/ICCUBEA.2016.7860040>
- Su, S., & Wang, J. (2023). Machine learning prediction of contents of oxygenated components in bio-oil using extreme gradient boosting method under different pyrolysis conditions. *Bioresource Technology*, 379, 129040. <https://doi.org/10.1016/j.biortech.2023.129040>
- Surendra, K. C., Angelidaki, I., & Khanal, S. K. (2022). Bioconversion of waste-to-resources (BWR-2021): Valorization of industrial and agro-wastes to fuel, feed, fertilizer, and biobased products. *Bioresource Technology*, 347, 126739. <https://doi.org/10.1016/j.biortech.2022.126739>
- Suriapparao, D. V., & Vinu, R. (2018). Effects of Biomass Particle Size on Slow Pyrolysis Kinetics and Fast Pyrolysis Product Distribution. *Waste and Biomass Valorization*, 9(3), 465–477. <https://doi.org/10.1007/s12649-016-9815-7>
- Tang, Q., Chen, Y., Yang, H., Liu, M., Xiao, H., Wu, Z., Chen, H., & Naqvi, S. R. (2020). Prediction of Bio-oil Yield and Hydrogen Contents Based on Machine Learning Method: Effect of Biomass Compositions and Pyrolysis Conditions. *Energy & Fuels*, 34(9), 11050–11060. <https://doi.org/10.1021/acs.energyfuels.0c01893>
- Taşar, Ş. (2022). Estimation of pyrolysis liquid product yield and its hydrogen content for biomass resources by combined evaluation of pyrolysis conditions with proximate-ultimate analysis data: A machine learning application. *Journal of Analytical and Applied Pyrolysis*, 165, 105546. <https://doi.org/10.1016/j.jaap.2022.105546>
- Trinh, T. N., Jensen, P. A., Dam-Johansen, K., Knudsen, N. O., & Sørensen, H. R. (2013). Influence of the Pyrolysis Temperature on Sewage Sludge Product Distribution, Bio-Oil, and Char Properties. *Energy & Fuels*, 27(3), 1419–1427. <https://doi.org/10.1021/ef301944r>
- Varma, A. K., Shankar, R., & Mondal, P. (2018). A Review on Pyrolysis of Biomass and the Impacts of Operating Conditions on Product Yield, Quality, and Upgradation. In *Recent Advancements in Biofuels and Bioenergy Utilization* (pp. 227–259). Springer Singapore. https://doi.org/10.1007/978-981-13-1307-3_10
- Velvizhi, G., Balakumar, K., Shetti, N. P., Ahmad, E., Kishore Pant, K., & Aminabhavi, T. M. (2022). Integrated biorefinery processes for conversion of lignocellulosic biomass to value added materials: Paving a path towards circular economy. *Bioresource Technology*, 343, 126151. <https://doi.org/10.1016/j.biortech.2021.126151>
- Venderbosch, R., & Prins, W. (2010). Fast pyrolysis technology development. *Biofuels, Bioproducts and Biorefining*, 4(2), 178–208. <https://doi.org/10.1002/bbb.205>
- Wang, L., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal*, 4(3), 212–219. <https://doi.org/10.1016/j.cj.2016.01.008>
- Wurzer, C., & Mašek, O. (2021). Feedstock doping using iron rich waste increases the pyrolysis gas yield and adsorption performance of magnetic biochar for emerging contaminants. *Bioresource Technology*, 321, 124473. <https://doi.org/10.1016/j.biortech.2020.124473>
- Xie, Q., Wang, G., Peng, Z., & Lian, Y. (2018). Machine Learning Methods for Real-Time Blood Pressure

- Measurement Based on Photoplethysmography. *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, 1–5. <https://doi.org/10.1109/ICDSP.2018.8631690>
- Xing, J., Luo, K., Wang, H., Gao, Z., & Fan, J. (2019). A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. *Energy*, *188*, 116077. <https://doi.org/10.1016/j.energy.2019.116077>
- Xiong, Z., Guo, J., Chaiwat, W., Deng, W., Hu, X., Han, H., Chen, Y., Xu, K., Su, S., Hu, S., Wang, Y., & Xiang, J. (2020). Assessing the chemical composition of heavy components in bio-oils from the pyrolysis of cellulose, hemicellulose and lignin at slow and fast heating rates. *Fuel Processing Technology*, *199*, 106299. <https://doi.org/10.1016/j.fuproc.2019.106299>
- Yang, K., Wu, K., & Zhang, H. (2022a). Machine learning prediction of the yield and oxygen content of bio-oil via biomass characteristics and pyrolysis conditions. *Energy*, *254*, 124320. <https://doi.org/10.1016/j.energy.2022.124320>
- Yang, Q., Mašek, O., Zhao, L., Nan, H., Yu, S., Yin, J., Li, Z., & Cao, X. (2021). Country-level potential of carbon sequestration and environmental benefits by utilizing crop residues for biochar implementation. *Applied Energy*, *282*, 116275. <https://doi.org/10.1016/j.apenergy.2020.116275>
- Yang, Y., Shahbeik, H., Shafizadeh, A., Masoudnia, N., Rafiee, S., Zhang, Y., Pan, J., Tabatabaei, M., & Aghbashlo, M. (2022b). Biomass microwave pyrolysis characterization by machine learning for sustainable rural biorefineries. *Renewable Energy*, *201*, 70–86. <https://doi.org/10.1016/j.renene.2022.11.028>
- Yang, Y., Shahbeik, H., Shafizadeh, A., Rafiee, S., Hafezi, A., Du, X., Pan, J., Tabatabaei, M., & Aghbashlo, M. (2023). Predicting municipal solid waste gasification using machine learning: A step toward sustainable regional planning. *Energy*, *278*, 127881. <https://doi.org/https://doi.org/10.1016/j.energy.2023.127881>
- Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*, 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Zeng, X., Shao, H., Pan, R., Wang, B., Deng, Q., Zhang, C., & Suo, T. (2022). Real-time damage analysis of 2D C/SiC composite based on spectral characters of acoustic emission signals using pattern recognition. *Acta Mechanica Sinica*, *38*(10), 422177. <https://doi.org/10.1007/s10409-022-22177-x>
- Zhang, T., Cao, D., Feng, X., Zhu, J., Lu, X., Mu, L., & Qian, H. (2022). Machine learning prediction of bio-oil characteristics quantitatively relating to biomass compositions and pyrolysis conditions. *Fuel*, *312*, 122812. <https://doi.org/10.1016/j.fuel.2021.122812>
- Zhang, W., Wu, C., Zhong, H., Li, Y., & Wang, L. (2021). Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geoscience Frontiers*, *12*(1), 469–477. <https://doi.org/10.1016/j.gsf.2020.03.007>
- Zhou, S., Tang, S., Li, G., Xin, S., Huang, F., Liu, X., Mi, T., Huang, K., & Zeng, L. (2023). Catalytic fast pyrolysis of herbal medicine wastes over zeolite catalyst for aromatic hydrocarbons production. *Fuel*, *333*, 126311. <https://doi.org/10.1016/j.fuel.2022.126311>



Using machine learning to predict the production and quality of bio-oil from pyrolysis biomass

EXTENDED ABSTRACT

Introduction

Reducing the reserves of fossil energy sources serves as a warning sign for humanity. On the other hand, the increasing consumption of fossil fuels has led to significant environmental problems, such as global warming. These issues make the replacement of renewable energy sources with fossil fuels inevitable. Among various renewable energy sources, biomass is a reliable and sustainable resource. Thermochemical conversions of biomass are a promising method for converting raw biomass into liquid (bio-oil), solid (bio-char), and gas (biogas) fuels suitable for modern life.

Biomass pyrolysis

Biomass pyrolysis is a process that involves heating organic matter in the absence of oxygen to produce bio-oil. The biofuel produced can be refined into transportation fuels or used as feedstock for chemical production. There are several different types of pyrolysis. Flash pyrolysis involves rapidly heating biomass to high temperatures without oxygen, forming a high-quality bio-oil. Slow pyrolysis, on the other hand, involves heating biomass at lower temperatures for longer periods, producing more biochar and less bio-oil. The pyrolysis process is complex and influenced by numerous variables, such as feedstock properties, heating rate, temperature, and residence time. Modeling can be a promising strategy to optimize the process parameters and improve product yield and quality.

Machine learning modeling

Machine learning is a field of artificial intelligence that uses algorithms to enable computers to learn from data and make predictions or decisions without being explicitly programmed. The process involves feeding data into a machine learning model, which then uses statistical analysis to identify patterns and relationships in the data. These patterns are used to make predictions or decisions about new, unseen data.

Supervised machine learning

Supervised machine learning is a type of machine learning where the algorithm is trained on labeled data, meaning the data has already been classified or grouped into different categories. The goal is to create a model to accurately predict the correct category or label for new, unseen data. There are two main subcategories of supervised machine learning: classification and regression. Classification involves predicting categorical labels, such as whether an email is spam or not. Regression, on the other hand, involves predicting a continuous numerical value. Several popular algorithms are used in supervised machine learning, including multi-linear regression, bagging regression decision trees, and random forest logistic regression.

Application of machine learning in the prediction of bio yield and quality prediction

Various studies have considered several independent inputs for modeling the pyrolysis process with machine learning methods. Regarding biomass composition, parameters such as the carbon content, hydrogen, oxygen, nitrogen, and sulfur of biomass, moisture content, ash content, and the amount of fixed carbon are considered effective parameters on the pyrolysis process in independent inputs. Operational input features include operating temperature, heating rate, reaction time, particle size, and input carrier fluid rate. Bio-oil yield, calorific value, carbon-to-hydrogen or carbon-to-oxygen ratio, sulfur and nitrogen in bio-oil, acidity, acid content, and various amounts of organic compounds have been considered dependence output parameters.

Application of parameter importance analysis in machine learning for predicting the production process and quality of bio-oil.

Although machine learning techniques can model complex phenomena, the nature of machine learning techniques, like a black box, makes interpreting results and identifying influential factors on the model very challenging. Therefore, the results of each machine learning model should be explained using advanced tools. These methods, such as Shapley Additive Explanations, provide a way to measure the impact of each independent input parameter of the pyrolysis process on the dependent response in the output of the machine learning models.

Results and Future Trends

Developing and expanding machine learning models for more complex processes and considering other influential parameters, such as catalyst properties on the biomass catalytic pyrolysis process, is possible. For instance, catalyst type, the amount used, and its structural and chemical properties greatly affect the product performance and quality improvement of bio-oil. Therefore, providing solutions in the future that can quantitatively incorporate important catalyst features as independent parameters during the modeling process

and accurately model the effects of these factors on the outputs would be highly significant. Additionally, the use of machine learning modeling in semi-industrial and industrial systems for biomass pyrolysis can be investigated to improve the performance of these units.