

Article

# Machine Learning Assisted Chemical Process Parameter Mapping on Lignin Hydrogenolysis

Yin Liu <sup>1</sup>, Shuo Cheng <sup>2</sup> and Jeffrey Scott Cross <sup>1,2,\*</sup> 

<sup>1</sup> Department of Materials Science and Engineering, School of Materials and Chemical Technology, Tokyo Institute of Technology, Tokyo 152-8550, Japan

<sup>2</sup> Department of Transdisciplinary Science and Engineering, School of Environment and Society, Tokyo Institute of Technology, Tokyo 152-8550, Japan

\* Correspondence: cross.j.aa@m.titech.ac.jp

**Abstract:** Lignin depolymerization has been studied for decades to produce carbon-neutral chemicals/biofuels and biopolymers. Among different chemical reaction pathways, catalytic hydrogenolysis favors reactions under relatively mild conditions, while its yield of bio-oil and high-value aromatic products is relatively high. In this study, the influence of reaction parameters on lignin hydrogenolysis are discussed by chemical process parameter mapping and modeled using three different machine learning algorithms based upon literature experimental data. The best R2 scores for solid residue and aromatic yield were 0.92 and 0.88 for xgboost, respectively. The parameter importance was examined, and it was observed that lignin-to-solvent ratio and average pore size have a larger impact on lignin hydrogenolysis results. Finally, the optimal conditions of lignin hydrogenolysis were predicted by chemical process parameter mapping using the best-fit machine learning model, which indicates that further process improvements can potentially generate higher yields in industrial applications.

**Keywords:** lignin hydrogenolysis; machine learning; LightGBM; XGBoost; CatBoost; chemical process parameter mapping



**Citation:** Liu, Y.; Cheng, S.; Cross, J.S. Machine Learning Assisted Chemical Process Parameter Mapping on Lignin Hydrogenolysis. *Energies* **2023**, *16*, 256. <https://doi.org/10.3390/en16010256>

Academic Editor: Valentina Colla

Received: 29 November 2022

Revised: 19 December 2022

Accepted: 23 December 2022

Published: 26 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

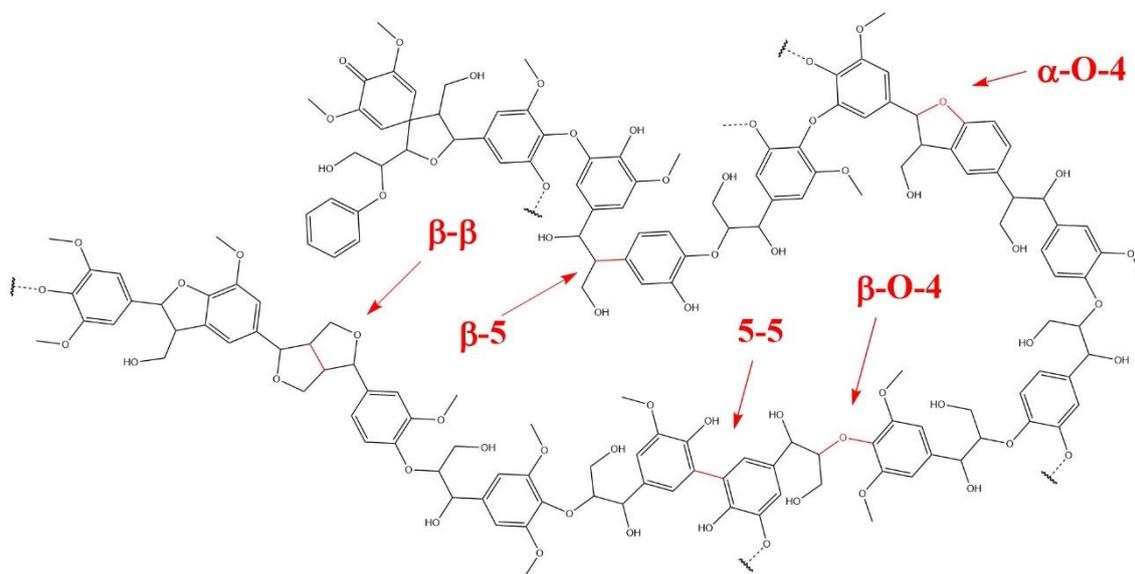
## 1. Introduction

According to the Organization for Economic Co-operation and Development (OECD) energy data, fossil fuels made up 77.8% of the global energy supply in 2020 [1] and also produce a large majority of global greenhouse gas emissions (GHG). Biomass is a source of carbon-neutral feedstocks and can potentially become a direct alternative substitute for fossil fuels if cost-effective processes can be developed. By replacing fossil fuels with bio-oil produced from biomass, the global reliance on the burning of fossil fuels can gradually be reduced.

Lignocellulose is the most promising type of biomass due to its low costs, abundance, and it also being non-edible, for example, agricultural residues [2], terrestrial plants, and recycled paper residues [3]. Lignocellulose consists of carbohydrate polymers (cellulose, hemicellulose) and an aromatic polymer, lignin, the second most abundant yet most underused component of lignocellulose [4]. Cellulose and hemicellulose depolymerization over heterogeneous catalysts has been well developed, and further efforts in this field should be directed towards its commercialization [5]. However, the depolymerization of lignin still needs further research and optimization due to the diverse chemical components and complicated structure of lignin.

Lignin has been defined as a macromolecular polymer, composed of three main building blocks: p-coumaryl alcohol (p-hydroxyphenyl, H), coniferyl alcohol (guaiacyl, G), and sinapyl alcohol (syringyl, S). All these monolignols include the phenyl group, and therefore their conjugated  $\pi$ -system causes the formation of various inter-unit linkages [6]. Among them, the most abundant linkage types are  $\beta$ -O-4 linkage (around 45–48%) and  $\alpha$ -O-4 linkage. These ether bonds are the easiest to cleave and where various depolymerization

methods are aimed. Other than regular carbohydrate polymers, there exists a bountiful number of C-C bonds, for example,  $\beta$ - $\beta$ , 5-5, and  $\beta$ -5 (Figure 1).



**Figure 1.** Main linkages in suppositional lignin structure.

Compared to other lignin depolymerization methods, lignin hydrogenolysis uses mild conditions, including lower temperature, less time, neutral solution, and results in better yields of bio-oil and small aromatic products. While mild conditions are more economic to operate and allow hydrogenolysis to have a higher potential for industrial application [7].

Based upon a literature review of lignin hydrogenolysis [7], the reaction environment plays a huge role in the outcome of the reaction, while the optimization of catalyst is currently a topic of heated discussion. Homogeneous catalyst may allow for greater potential for catalysts to reach the lignin or degraded lignin fragments; the difficulty of obtaining the various products that are separate from the catalyst in the reaction mixture hampered their development [8]. On the other hand, there are many characteristics to consider in heterogeneous catalysts; among them, catalyst acidity and choice of active phase are known to be especially important, as both of them positively impact the reaction [9]. Accordingly, either metal directly impregnated on an acidic catalyst support [10] or combined metal-supported neutral catalyst and acidic catalyst in a one-pot reaction [11] has been studied and proved to be effective.

Currently, what is lacking in the literature is a study that investigates the effect of multiple catalyst features on lignin hydrogenolysis. One possible pathway is to study the performance of multi-functional catalysts via model reactions, which can be further directly applied to lignin hydrogenolysis. Studies regarding the upgrading of lignin-derived products have been well developed recently, but there is still a lack of studies regarding the conversion of lignin to aromatics under mild conditions. In fact, in order to achieve a more efficient lignin hydrogenolysis, the selection of reaction conditions is somewhat arbitrary depending upon the desired products. A new approach to locate the optimized process parameters and conditions is to consider the use of chemical process parameter mapping by modeling already published data using machine learning algorithms. This method has the potential to grant greater insight into which process parameter or feature has the largest impact on target yields.

Chemical space mapping is a designed optimization procedure of generative topographic maps for the purpose of the simultaneous selection and dimensionality reduction of a possible descriptor, which has been widely used to predict the reaction products in recent studies. Kyrylo et al. successfully predicted the antiviral activities of a specific chemical compound based on generative topographic mapping [12]. Furkan et al. reported

successful machine learning modeling and prediction from the component of biomass to products with an artificial neural network (ANN) with 178 data points [13]. In this study, the reaction parameters instead of products of lignin hydrogenolysis were examined and modeled through the so-called chemical process parameter mapping to analyze how they impact the yield of char and aromatics. Notably, metal content along with catalyst acidity, measured by ammonium temperature programmed desorption, which were two important active sites in lignin hydrogenolysis, were also examined.

Machine learning for efficient parameter modeling is an effective tool. A particular type of machine learning model, which would be better applied to lignin hydrogenolysis results prediction, is Gradient Boosting Tree models, such as LightGBM [14], XgBoost [15], and CatBoost [16] models. In this study, lignin hydrogenolysis reaction conditions were examined first by considering the process parameter mapping of the reaction environment and then modeled using machine learning (LGBM, XgBoost, CatBoost) models to examine the catalyst utility in relationship to the product yields. Castro et al. reported a successful prediction of lignin depolymerization with random forest (RF) modeling [17]. It is found that the lignin depolymerization results are more dependent on features associated with solvolysis than with the catalysts. However, in this study, we will focus on analysis of a particular type of lignin depolymerization, which would allow the lignin to depolymerize at a relatively milder condition. The impact of each feature will be re-examined in this range.

## 2. Materials and Methods

### 2.1. Data Collection and Pre-Processing

Data from recently published papers and conferences articles, from 2017 to 2022, which reported lignin hydrogenolysis, were collected using the Google Scholar search engine. Keywords such as “lignin” and “hydrogenolysis” were used. To refine the papers that report the desiring data, “TPD”, “BET”, and “GC-MS” were added into the keywords searching bar, where “TPD” stands for ammonia *Temperature Programmed Desorption*, usually used for characterization of catalyst acidity, and “BET” stands for *Brunauer–Emmett–Teller*, a theory applied in N<sub>2</sub> adsorption-desorption, which is a characterization method for testing catalyst physical properties. GC-MS stands for *Gas Chromatography–Mass Spectrum*, an often-used characterization method for analyzing lignin reaction products. After refining, 731 results remain, most of which are about pyrolysis or reactions on a furfural-derived macromolecular polymer. As a result, a manual filtration was applied, and papers with ambiguous statements of measurements’ results and papers working on model compound reactions were also separated. In total, from 2021 July to 2022 July, the online research was done on google scholar, 25 papers were selected, and 263 data sets were analyzed in this study as shown in Table 1.

**Table 1.** Selected lignin hydrogenolysis papers comparison of catalyst and yields analyzed by ML.

Catalyst Support	Metal	Solid Yield	Aromatics Yield	Published Year	Ref.
CSA (carbonaceous solid acids)	none	0~12.40	9.11~32.83	2019	[18]
HZSM-5 (zeolite)	none	15.36~19.80	10.50~12.29	2018	[19]
H-beta (zeolite)	Ni	9.52~54.68	3.43~11.79	2018	[20]
PTA (Terephthalic acid)	Ni, Cs	0~21.61	12.10~20.28	2017	[21]
TiO <sub>2</sub> , ZrO <sub>2</sub>	Ni, Co	7.25~54.13	3.4~18.18	2019	[22]
Nb <sub>2</sub> O <sub>5</sub>	Ni, Re	2.38~27.01	9.93~30.4	2020	[23]
Zeolites (H-Beta, HZSM-5, MAS-7, MCM-41 SAPO-11)	Ni, Cu	0.04~27.01	0~20.10	2019	[24]
Active carbon, MgO, ZrO <sub>2</sub> , Al <sub>2</sub> O <sub>3</sub>	Ru	1.75~6.1	8.8~47.3	2018	[25]
MgO, AC	Pd	~4.32	~24.49	2019	[26]

**Table 1.** *Cont.*

Catalyst Support	Metal	Solid Yield	Aromatics Yield	Published Year	Ref.
de-aluminum H-beta (zeolite)	Co, Zn	8.60~51.63	2.39~12.21	2020	[27]
MCM-41, PTA	none	3.33~44.46	0.27~8.79	2020	[28]
Zeolites (H-FER, H-MOR, H-Beta)	none	29.0~41.0	1~3.4	2017	[29]
SBA-15 (mesoporous silica)	Ni, Ru	3.37~30.52	5.8~12.7	2018	[30]
Active carbon, MgO, ZrO <sub>2</sub> , Al <sub>2</sub> O <sub>3</sub>	Ni, Ru, Pd	4.2~42.8	7.44~35.49	2020	[31]
Cr <sub>2</sub> O <sub>3</sub>	Cu, Pd	25.31~56.16	2.37~11.24	2021	[32]
Nitrogen doped biochar	Ru	19.3~26.5	16.9~29.4	2022	[33]
Activate carbon	Ru	~56.1	~13.24	2022	[34]
Carbon nanotube	Ru	4.7~32.6	1.2~42.7	2022	[35]
Nitrogen-iron doped carbon nanotube	Ni	~60.5	~20.2	2022	[36]
Ga-doped ZSM-5 (zeolite)	Ru	~40.5	~26.76	2022	[37]
HY (zeolite)	Ni, Ru	5.23~18.3	5.5~20.2	2022	[38]
Iron dispersed HZSM-5 (zeolite)	Pd	1~39.06	8.37~27.93	2021	[39]
Al-SBA-15	Ni	15.25~20.17	17.87~21.56	2021	[40]
Activate carbon	none	6.61~56.68	18.76~40.4	2021	[41]
HZSM-5 (zeolite)	Ru	~48.7	~19.95	2021	[42]

Data related to reaction conditions, catalyst properties, and reaction results are tabulated. In total, 17 variables are converted into standardized units, as listed in Table 2, including reaction environments, for example, time and temperature, as well as catalyst properties, such as pore attributes and metal contents. Regarding the model outputs, we selected a range of aromatics (Supplementary Material S1) to evaluate the performance of each reaction system. Raw data table is provided as Supplementary Material S2.

**Table 2.** Variables and their abbreviations that applied in this research.

Variable Name	Abbreviation	Range	Unit
Reaction temperature	TEMP	160–400	°C
Reaction time	TIME	0.5–36	h
Hydrogen pressure	HPRE	0–4	MPa
Solvent-to-Reactor Ratio	SRR	0.16–0.6	-
Lignin-to-Solvent Ratio	LSR	1.75–50	-
Solvent of Water Ratio	SOWR	0.1667–1	-
Solvent of Alcohol Ratio	SOAR	0–1	-
Solvent of Organic Ratio	SOOR	0–1	-
Catalyst-to-Lignin Ratio	TCR	0.01–2	-
Average pore size	POSZ	0.43–20.4	nm
Average pore volume	PRVL	0.005–1.95	cm <sup>3</sup> /g
BET surface area	BETS	4.2–3349	m <sup>2</sup> /g
Total acidity	ACID	0–49	mmol NH <sub>3</sub> /g
Cobalt ratio in catalyst	MCOR	0–4.59	-
Nickel ratio in catalyst	MNIR	0–0.5	-
Ruthenium ratio in catalyst	MRUR	0–0.2	-
Palladium ratio in catalyst	MPDR	0–0.05	-

## 2.2. Machine Learning Algorithm and Evaluation Indicators

### 2.2.1. Algorithms

According to the previous work reported by Castro et al., the RF regression method predicts lignin depolymerization yields well [17]. Gradient boosting decision trees (GBDT) can provide a high degree of transparency and intelligibility. Here, three different algorithms are addressed and discussed: LightGBM, XgBoost and CatBoost, which are provided by Ke et al. (2017), Chen and Guestrin (2015), and Prokhorenkova et al. (2019). All codes were written in Python and the raw code is provided as Supplementary Material S4.

LightGBM provides one-side sampling, which can avoid going down branches that are of less importance. It also uses Exclusive Feature Bundling method to reduce feature sparsity, making it possible to decrease the model's training duration. The drawback of LightGBM is that it can overfit when applied to smaller datasets. On the other hand, XgBoost provides full-side parallel tree expansion. XGBoost can quickly consume the available memory and potentially crash a computer when dealing with more massive and complex datasets. The CatBoost puts greater emphasis on categorical features, which combats the issue of exponential feature combination growth by using an efficient method at every new split of the current tree.

The datasets were randomly separated into train sets and test sets for model training (ratio of 7:3 (train to test)). Hyperparameters of all three models are tuned separately through Randomized Search Cross Validation (RandomizedSearchCV) among the selected range as shown in Table S3-1 in Supplementary information S3 using 5-fold cross validation, applying *mean absolute error* for evaluation during the tuning.

### 2.2.2. Evaluation Indicators

Several functions from the library, such as R-squared ( $R^2$ ), mean-squared-error (MSE), and Pearson correlation coefficient (PCC), are applied to evaluate and modify the machine learning model, whose equations are shown below, (1), (2) and (3).

$$R^2 = 1 - \frac{\sum_i (y_p - y_i)^2}{\sum_i (y_p - \bar{y}_i)^2} \quad (1)$$

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

$$PCC = \frac{\sum (y_i - \bar{y}_i)(y_p - \bar{y}_p)}{\sqrt{\sum (y_i - \bar{y}_i)^2 \sum (y_p - \bar{y}_p)^2}} \quad (3)$$

where  $y_p$  denotes the predicted value,  $y_i$  represents the experiment value, and  $n$  represents the number of test samples.

$R^2$  score and prediction curve is obtained for comparison. The importance of the features was examined via an open access method named SHAP (SHapley Additive ex-Planations) using the TreeSHAP [43] method. It defines the value function using the conditional expectation to estimate effects on the model output value. Both linear fitting and polynomial (5th) fitting were performed via OriginLab program.

### 2.2.3. Chemical Parameter Mapping

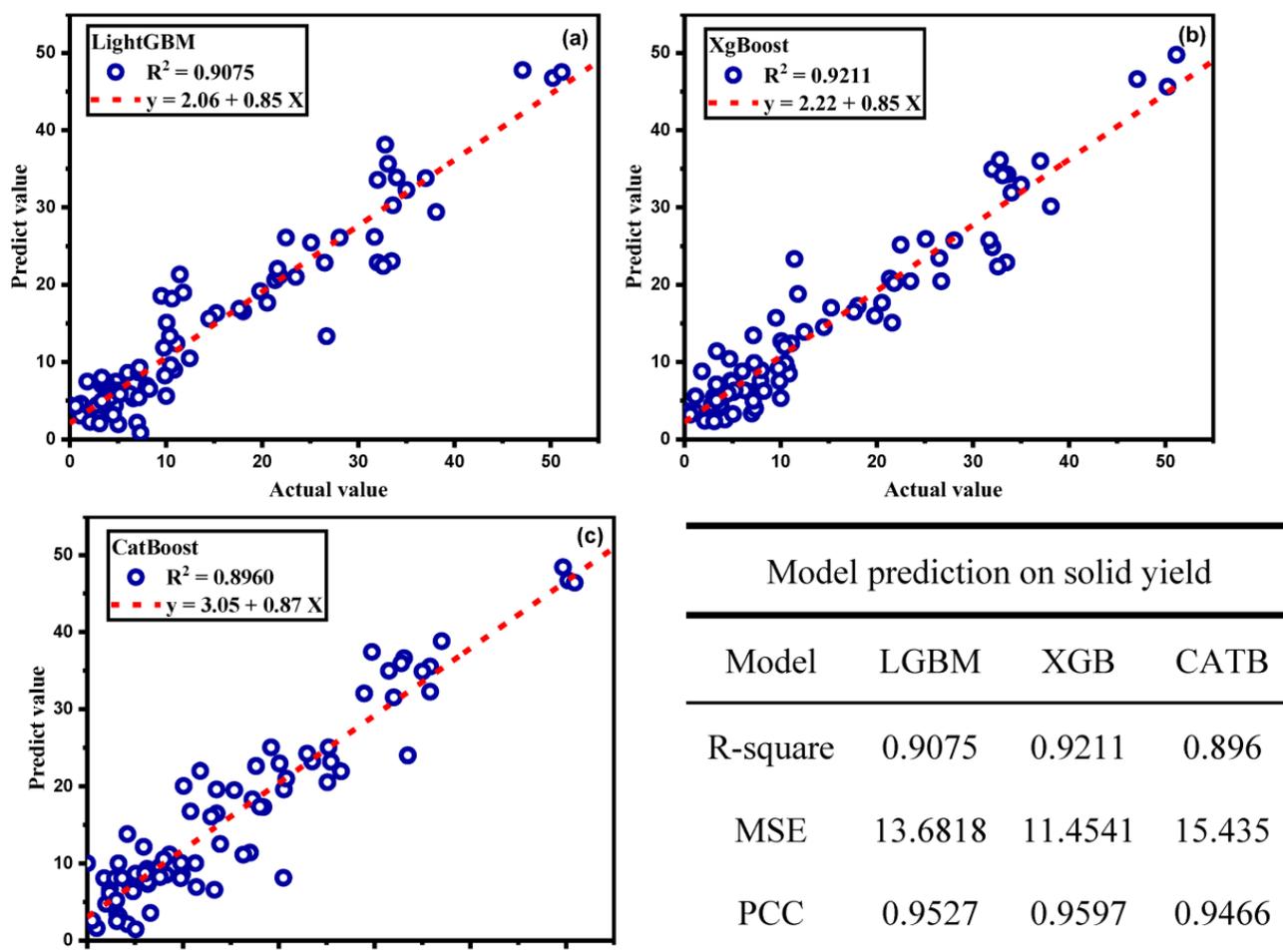
The mapping was performed separately with the best fit trained model via machine learning. For each parameter, a total of 200 datapoints were simulated for investigating the exact numerical influence on the solid yields results. A reference data point was selected from the training data sets. The simulating datapoints were then created based on the selected datapoint. The value of the selected feature was varied in the range of Table 2, while all other features' values remain same as the reference datapoint

## 3. Results and Discussion

### 3.1. Prediction Curve and Analysis of Solid Residue

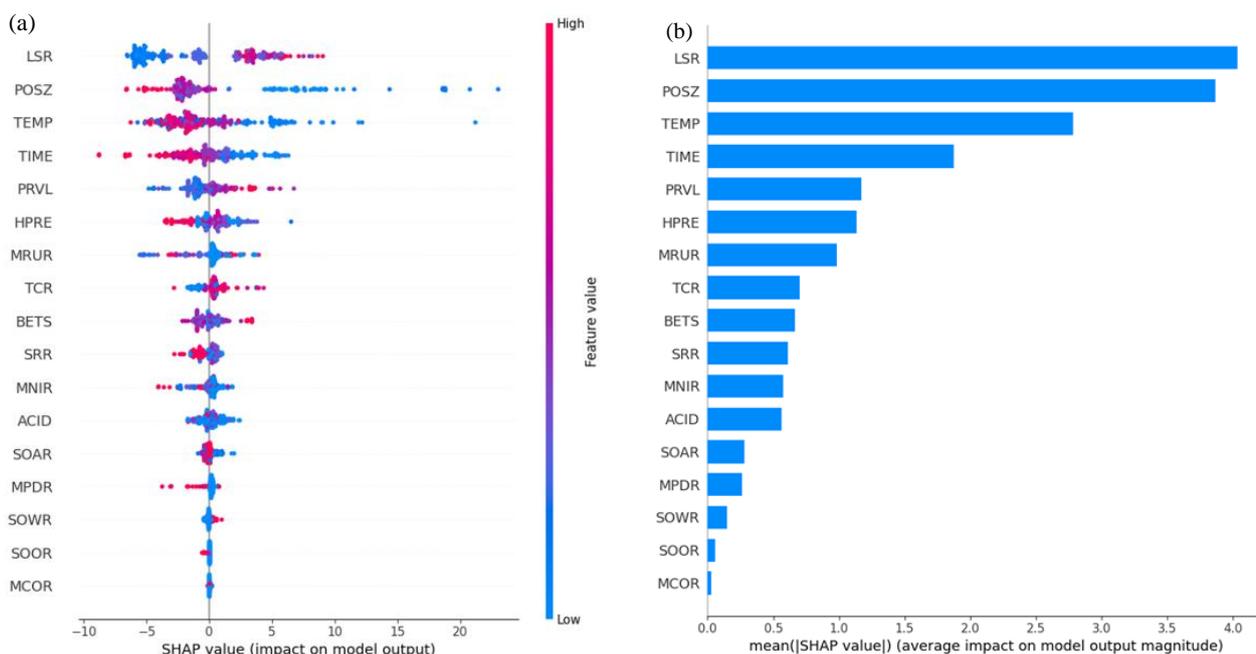
In terms of reliability of solid yield data, some papers report a range of results from repeated experiments, and the average of the results under the same conditions was selected. The lignin hydrogenolysis also depends on several other features that were ignored in this study, for example, the lignin source (chemical composition or molecular weight). All lignin literature used in this study mainly used alkaline lignin or kraft lignin. Alkaline lignin is reported to have a slightly higher molecular weight and more glycerol end-groups [44].

After tuning, all three models showed high R-square value against training sets: 0.97 for LightGBM; 0.99 for XgBoost; 0.98 for CatBoost. However, when applied to the test sets, all the evaluation scores decreased to some degree. This is mainly caused by the data variation in the published journals. After all, the XGB model achieved an R-square score of 0.92 and a PCC score of 0.96, showing its excellent performance of prediction against a solid yield of lignin hydrogenolysis (Figure 2).



**Figure 2.** Solid yield prediction curve and prediction score of (a) LightGBM, (b) XgBoost, (c) CatBoost and table of machine learning evaluation results.

The feature importance was then examined via the SHAP method (Figure 3). Lignin-to-solvent ratio has the highest impact among all the features, while the size of the catalyst pore showed a higher impact than other pore attributes. The solubility of lignin in most solvents is mostly limited; thus, the reaction would remain a heterogeneous state and lignin can continuously dissolve during the reaction. In the meantime, the solvent amount can affect the amount of achievable hydrogen. However, the type of solvent is, relatively speaking, of less value. To our knowledge, the solubility of lignin is better in alcohols than in water and most organic solvents, which collapse with the result that solvent of alcohol ratio has a higher impact than the other two types of solvents. Notably, from all models' prediction results, nickel and ruthenium showed higher effects on the models' output than other metals (SHAP value and feature importance figures of LightGBM model and CatBoost model predictions on solid yield are shown in Figures S3-1 and S3-3 in Supplementary Material S3).



**Figure 3.** SHAP values (a) and feature importance (b) of XgBoost prediction on solid yield.

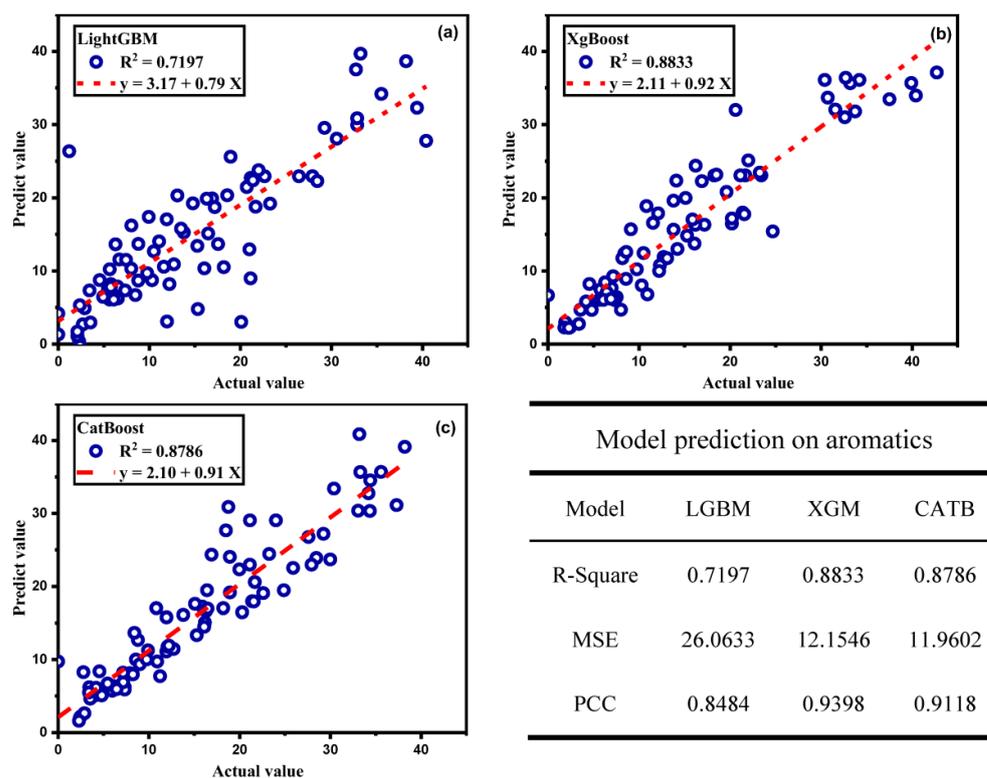
The pore size has a higher importance than other pore attributes. The termination of activating free radicals in the middle of a reaction is essential for preventing the formation of a stable carbon–carbon bond, which further causes the production of highly stabilized char. A catalyst with suitable pore size ranges is desirable in order to terminate and stabilize bio-mediated reactions more easily, while, in the meantime, it allows lignin to reach the active sites in the catalysts. As a result, the catalyst pore size can greatly affect the solid yield.

### 3.2. Prediction Curve and Analysis of Aromatic Products

Although the yield of solid residue and aromatics have a considerable relationship, the decrease in solid residue does not exactly represent the increase in aromatics, as the residue could also transform into complex oligomers.

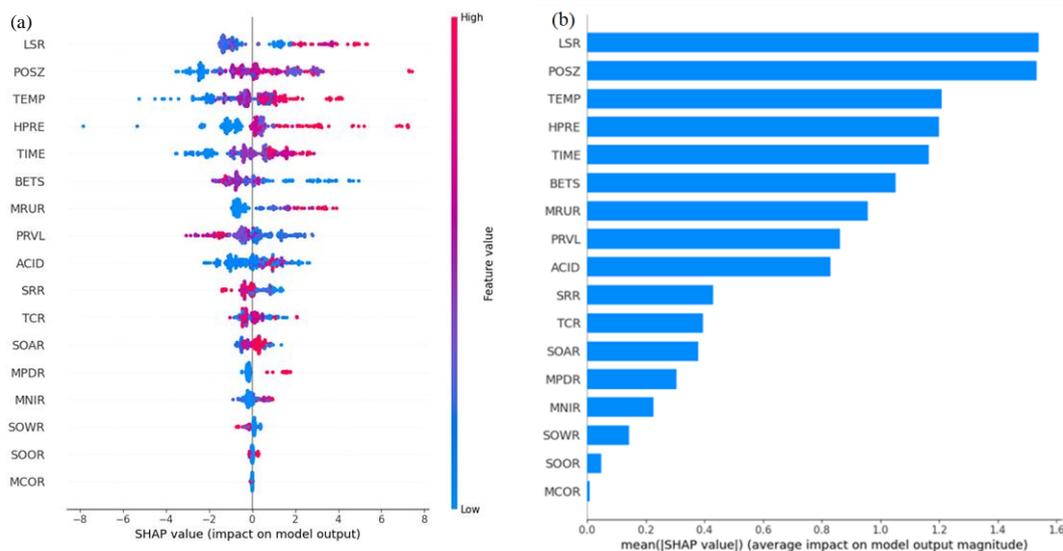
In general, the prediction of aromatic products is worse than the prediction of solid residue, which is mainly due to the variation in the reported values in the data. Differences between GC-MS analysis results and the manual selection of GC peaks can lead to differences in the yield of the aromatic products varying slightly. When the degree of error exceeds the machine learning processing capacity, the model's accuracy decreases correspondingly.

The R-square score of all three models during training were over 0.95: 0.97 for LightGBM; 0.95 for Xgboost; 0.97 for CatBoost. As we can observe from (Figure 4), the R-square score of LightGBM was the lowest at 0.72. When dealing with smaller datasets containing plenty noises, the accuracy of LightGBM is insufficient. R-square of XgBoost and CatBoost remains around 0.88. The slope of the best fit prediction curve also reached 0.9, indicating their good performance in aromatic prediction. Combining PCC score and the prediction curve, the performance of XgBoost is slightly better than CatBoost against aromatic yield prediction.



**Figure 4.** Aromatics yield prediction curve and prediction score of (a) LightGBM, (b) XgBoost, (c) CatBoost and table of machine learning evaluation results.

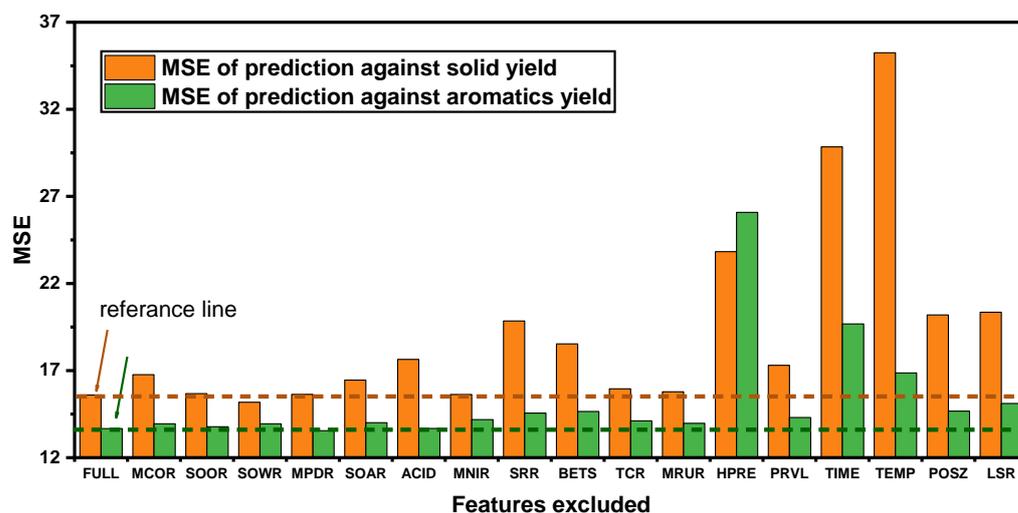
The feature importance was also examined via the SHAP method (Figure 5, Figures S3-2 and S3-4). The result of aromatic prediction was similar to the solid yield prediction results, where lignin-to-solvent ratio and pore size have higher impacts on results. However, compared with the solid yield prediction, the importance of hydrogen pressure is higher. The lignin hydrogenolysis can be generally divided into two steps, from lignin to intermediated products and from intermediated products to benzene derivatives or further alkanes. The second step requires a higher consumption of hydrogen, which explains the larger importance in aromatic prediction than in solid yield prediction.



**Figure 5.** SHAP value (a) and feature importance (b) of XgBoost prediction on aromatics.

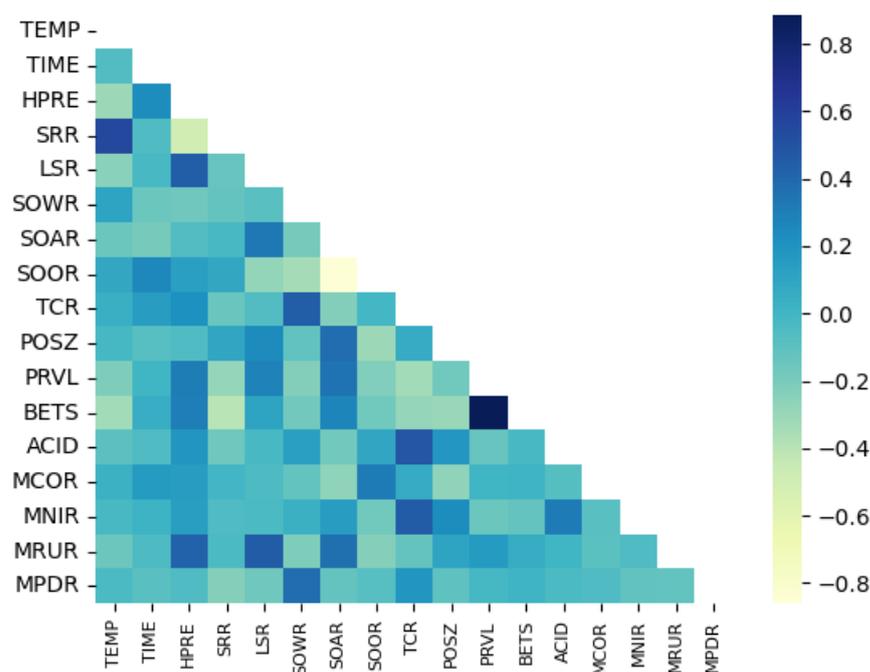
### 3.3. Contributions of Feature to the Machine Learning Performances

The contributions of each feature to the machine learning performance are not only examined via SHAP value function, but also studied via direct measurements of MSE change before and after excluding the selected feature. As observed in Figure 6, the mean square error of the prediction results and actual results all increased except for SOWR to solid yield and MPDR to aromatics yield. Overall, these features contribute positively to the performance of machine learning models.



**Figure 6.** Comparisons of mean square error of machine learning prediction before and after excluding each selected feature.

Subsequently, for a better understanding of the reaction results under mild conditions, a Pearson correlation coefficient matrix of these parameters was developed and is shown in Figure 7. It can be observed that, besides the correlation between pore volume and BET surface area, most of these parameters are poorly correlated to each other. In other words, the parameters of reaction conditions are independent of each other.



**Figure 7.** Pearson correlation coefficient matrix of all features.

### 3.4. Reaction Parameter Mapping

The chemical process mapping of the reaction parameter is simplified by a step-by-step mapping strategy in this study. In this study, a data point near the mean average of the output value and close to the prediction curve is selected as the reference data point. Afterwards, a range of designed conditions were set up within the range of the training datasets based on the selected reference data point, and the predictions of the designed conditions were tested via the previous trained XgBoost model.

The hydrothermal treatment of lignin-derived intermediated products has been well developed. One of the main targets of this research is to propose a suitable range of reaction conditions of direct lignin hydrogenolysis. In the previous section, we observed that LSR (lignin-to-solvent ratio) and POSZ (average pore size) had the largest impact on lignin hydrogenolysis. For the catalyst, the catalyst active sites are important, such as the metal type and its acidity, e.g., MNIR (nickel ratio in catalyst) and MRUR (ruthenium ratio in catalyst) showed the highest relevance when modeling using machine learning.

Consequently, the actual numerical effect of LSR, POSZ, MNIR and MRUR to solid yield of lignin hydrogenolysis are discussed via chemical parameter mapping. The predicted results are fitted using a third-degree polynomial for investigation.

As is shown in Figures 8 and 9, in order to achieve a lower yield of solid residue in lignin hydrogenolysis, a lignin-to-solvent ratio around 6 is optimal. A higher lignin-to-solvent ratio would lead to the insufficient reaction of lignin, and a lower lignin-to-solvent ratio would cause a lower diffusion of lignin in the catalyst system. The predicted curve of solid yield with respect to pore size showed three referable ranges, which are less than 2 nm, around 7 and 16 nm, which corresponds to the pore size of microporous, mesoporous and microporous materials. The optimum metal content in lignin hydrogenolysis is predicted to be less than 5% nickel and around 4% ruthenium.

Combining all the prediction results, it is proposed that, based on the data referred to in this study, a bimetallic catalyst with the ratio of less than 5% nickel, around 4% ruthenium and a lignin-to-solvent ratio of around 6 could perform the best result under a mild condition, which remarkably matches Feng, L's recent experimental work [45].

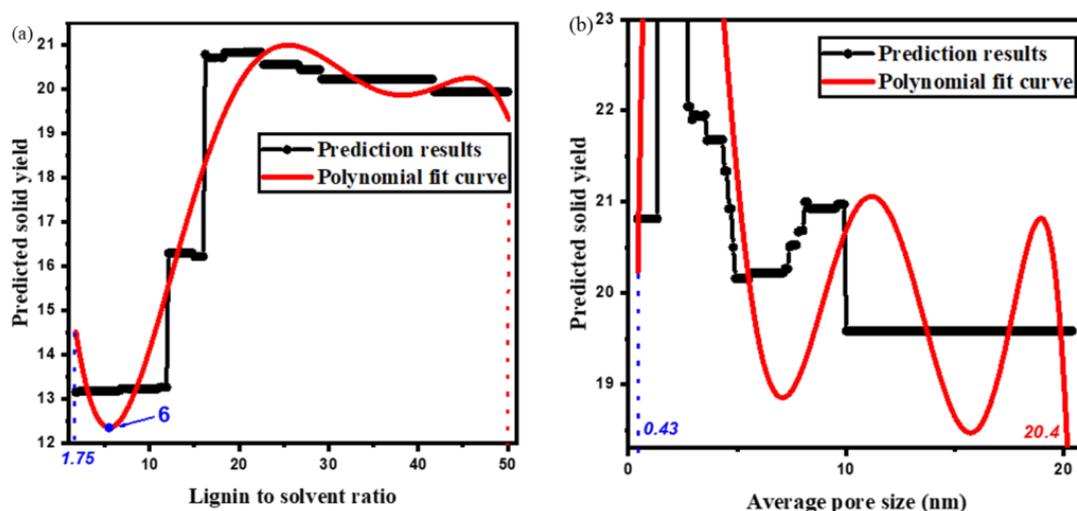


Figure 8. Predicted solid yield of designed lignin-to-solvent ratio (a) and average pore size (b).

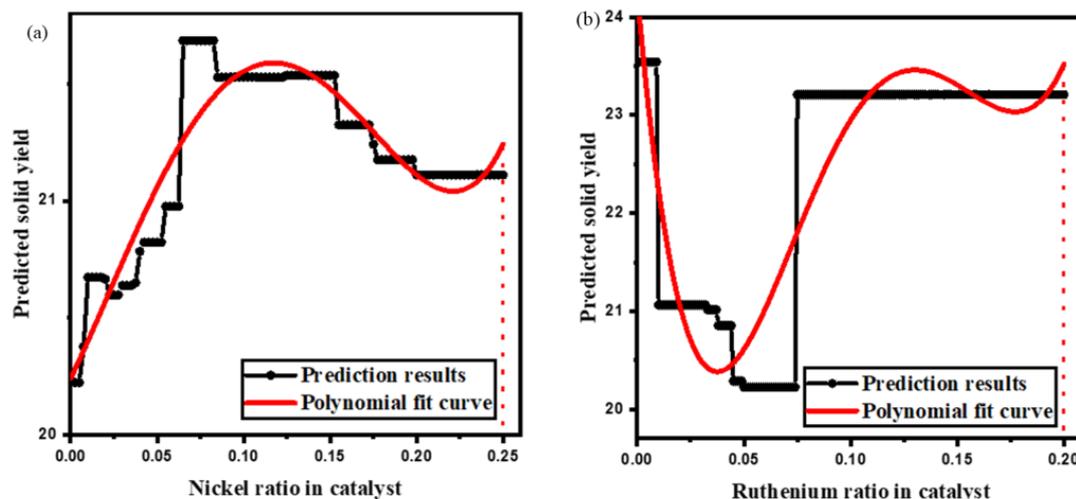


Figure 9. Predicted solid yield of designed nickel (a) and ruthenium (b) ratio in catalyst.

#### 4. Conclusions

In this study, three different ML models were tested on catalyzed lignin hydrogenolysis reaction conditions to predict product yields. XGBoost shows the best prediction abilities, resulting as R2 scores of 0.92 and 0.88 against solid residue and aromatics yields, respectively. Lignin-to-solvent ratio and average catalyst pore size are predicted to be the two most influential features to machine learning modeling. Among all active sites, including catalyst acidity and different kinds of metal, nickel and ruthenium play a significant role. With the assistance of a newly proposed chemical process parameter mapping scheme, near optimal conditions are proposed. By using machine learning and chemical process parameter mapping, further catalytic process optimization on lignin/bio-polymer reactions resulting in increased product yields and economic process are expected.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/en16010256/s1>. Supplementary material S1. List of aromatics in this study. Supplementary material S2. Table of datasets applied in this study. Supplementary material S3. Extra Tables and Figures, including Table S3-1: Machine learning models' hyperparameter tuning range; Figure S3-1: SHAP value (left) and feature importance (right) of LightGBM prediction on solid; Figure S3-2: SHAP value (a) and feature importance (b) of LightGBM prediction on aromatics; Figure S3-3: SHAP value (a) and feature importance (b) of CatBoost prediction on solid yield; Figure S3-4: SHAP value (a) and feature importance (b) of CatBoost prediction on aromatics. Supplementary material S4. Raw code of this study in the form of python file.

**Author Contributions:** Y.L.: Conceptualization, Investigation, Methodology, Writing—original draft. S.C.: Writing—review and editing. J.S.C.: Supervision, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data can be found in the supplementary materials.

**Acknowledgments:** Special thanks to Tokyo Tech Cross lab doctoral student Abraham Castro Garcia for his support regarding the python coding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. IEA. *Key World Energy Statistics 2021*; IEA: Paris, France, 2021. Available online: <https://www.iea.org/reports/key-world-energy-statistics-2021> (accessed on 1 July 2022).
2. Clementine, C.L.; Chen, M.; Fennell, P.S.; Hallett, J.P. Efficient Fractionation of Lignin- and Ash-Rich Agricultural Residues Following Treatment with a Low-Cost Protic Ionic Liquid. *Front. Chem.* **2019**, *7*, 246. [CrossRef]
3. Tofani, G.; Cornet, I.; Tavernier, S. Separation and recovery of lignin and hydrocarbon derivatives from cardboard. *Biomass Conv. Bioref.* **2022**, *12*, 3409–3424. [CrossRef]
4. Anderson, W.F.; Akin, D.E. Structural and Chemical Properties of Grass Lignocelluloses Related to Conversion for Biofuels. *J. Ind. Microbiol. Biotechnol.* **2008**, *35*, 355–366. [CrossRef] [PubMed]
5. Shrotri, A.; Kobayashi, H.; Fukuoka, A. Cellulose Depolymerization over Heterogeneous Catalysts. *Acc. Chem. Res.* **2018**, *51*, 761–768. [CrossRef]
6. De Jong, E.; Gosselink, R.J.A. Lignocellulose-Based Chemical Products. In *Bioenergy Research: Advances and Applications*; Elsevier: Amsterdam, The Netherlands, 2014; pp. 277–313.
7. Cheng, C.; Shen, D.; Gu, S.; Luo, K.H. State-of-the-Art Catalytic Hydrogenolysis of Lignin for the Production of Aromatic Chemicals. *Catal. Sci. Technol.* **2018**, *8*, 6275–6296. [CrossRef]
8. DeLucia, N.A.; Das, N.; Overa, S.; Paul, A.; Vannucci, A.K. Low Temperature Selective Hydrodeoxygenation of Model Lignin Monomers from a Homogeneous Palladium Catalyst. *Catal. Today* **2018**, *302*, 146–150. [CrossRef]
9. Kaur, P.; Singh, G.; Arya, S.K. Tandem catalytic approaches for lignin depolymerization: A review. *Biomass Convers. Biorefinery* **2022**, 1–12. [CrossRef]
10. Hu, L.; Wei, X.-Y.; Kang, Y.-H.; Guo, X.-H.; Xu, M.-L.; Zong, Z.-M. Mordenite-Supported Ruthenium Catalyst for Selective Hydrodeoxygenation of Lignin Model Compounds and Lignin-Derived Bio-Oil to Produce Cycloalkanes. *J. Energy Inst.* **2021**, *96*, 269–279. [CrossRef]
11. Wang, X.; Zhu, S.; Wang, S.; He, Y.; Liu, Y.; Wang, J.; Fan, W.; Lv, Y. Low Temperature Hydrodeoxygenation of Guaiacol into Cyclohexane over Ni/SiO<sub>2</sub> Catalyst Combined with H $\beta$  Zeolite. *RSC Adv.* **2019**, *9*, 3868–3876. [CrossRef]
12. Klimenko, K.; Marcou, G.; Horvath, D.; Varnek, A. Chemical Space Mapping and Structure–Activity Analysis of the ChEMBL Antiviral Compound Set. *J. Chem. Inf. Model.* **2016**, *56*, 1438–1454. [CrossRef]
13. Kartal, F.; Özveren, U. An improved machine learning approach to estimate hemicellulose, cellulose, and lignin in biomass. *Carbohydr. Polym. Technol. Appl.* **2021**, *2*, 2666–8939. [CrossRef]
14. Guolin, K.; Qi, M.; Thomas, F.; Taifeng, W.; Wei, C.; Weidong, M.; Qiwei, Y.; Tie-Yan, L. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Curran Associates, Inc.* 2017, p. 30. Available online: <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf> (accessed on 28 November 2022).
15. Suzuki, K.; Toyao, T.; Maeno, Z.; Takakusagi, S.; Shimizu, K.; Takigawa, I. Statistical Analysis and Discovery of Heterogeneous Catalysts Based on Machine Learning from Diverse Published Data. *ChemCatChem* **2019**, *11*, 4537–4547. [CrossRef]
16. Liudmila, P.; Gleb, G.; Aleksandr, V.; Anna V., D.; Andrey, G. CatBoost: Unbiased Boosting with Categorical Features. *Adv. Neural Inf. Process. Syst.* **2018**, 31. [CrossRef]
17. Castro Garcia, A.; Shuo, C.; Cross, J.S. Machine Learning Based Analysis of Reaction Phenomena in Catalytic Lignin Depolymerization. *Bioresour. Technol.* **2022**, *345*, 126503. [CrossRef] [PubMed]
18. Asawaworarit, P.; Daorattanachai, P.; Laosiripojana, W.; Sakdaronnarong, C.; Shotipruk, A.; Laosiripojana, N. Catalytic Depolymerization of Organosolv Lignin from Bagasse by Carbonaceous Solid Acids Derived from Hydrothermal of Lignocellulosic Compounds. *Chem. Eng. J.* **2019**, *356*, 461–471. [CrossRef]
19. Wang, X.; Du, B.; Pu, L.; Guo, Y.; Li, H.; Zhou, J. Effect of Particle Size of HZSM-5 Zeolite on the Catalytic Depolymerization of Organosolv Lignin to Phenols. *J. Anal. Appl. Pyrolysis* **2018**, *129*, 13–20. [CrossRef]
20. Li, W.; Dou, X.; Zhu, C.; Wang, J.; Chang, H.; Jameel, H.; Li, X. Production of Liquefied Fuel from Depolymerization of Kraft Lignin over a Novel Modified Nickel/H-Beta Catalyst. *Bioresour. Technol.* **2018**, *269*, 346–354. [CrossRef]
21. Shen, X.-J.; Wen, J.-L.; Huang, P.-L.; Zheng, K.; Wang, S.-F.; Liu, Q.-Y.; Charlton, A.; Sun, R.-C. Efficient and Product-Controlled Depolymerization of Lignin Oriented by Raney Ni Cooperated with Cs x H<sub>3</sub>-x PW12O<sub>40</sub>. *Bioenergy Res.* **2017**, *10*, 1155–1162. [CrossRef]
22. Zhu, C.; Dou, X.; Li, W.; Liu, X.; Li, Q.; Ma, J.; Liu, Q.; Ma, L. Efficient Depolymerization of Kraft Lignin to Liquid Fuels over an Amorphous Titanium-Zirconium Mixed Oxide Supported Partially Reduced Nickel-Cobalt Catalyst. *Bioresour. Technol.* **2019**, *284*, 293–301. [CrossRef]
23. Kong, L.; Zhang, L.; Gu, J.; Gou, L.; Xie, L.; Wang, Y.; Dai, L. Catalytic Hydrotreatment of Kraft Lignin into Aromatic Alcohols over Nickel-Rhenium Supported on Niobium Oxide Catalyst. *Bioresour. Technol.* **2020**, *299*, 122582. [CrossRef]
24. Kong, L.; Liu, C.; Gao, J.; Wang, Y.; Dai, L. Efficient and Controllable Alcoholysis of Kraft Lignin Catalyzed by Porous Zeolite-Supported Nickel-Copper Catalyst. *Bioresour. Technol.* **2019**, *276*, 310–317. [CrossRef] [PubMed]
25. Limarta, S.O.; Ha, J.-M.; Park, Y.-K.; Lee, H.; Suh, D.J.; Jae, J. Efficient Depolymerization of Lignin in Supercritical Ethanol by a Combination of Metal and Base Catalysts. *J. Ind. Eng. Chem.* **2018**, *57*, 45–54. [CrossRef]
26. Lv, W.; Zhu, Y.; Liu, J.; Wang, C.; Xu, Y.; Zhang, Q.; Chen, G.; Ma, L. Modifying MgO with Carbon for Valorization of Lignin to Aromatics. *ACS Sustain. Chem. Eng.* **2019**, *7*, 5751–5763. [CrossRef]

27. Dou, X.; Jiang, X.; Li, W.; Zhu, C.; Liu, Q.; Lu, Q.; Zheng, X.; Chang, H.; Jameel, H. Highly Efficient Conversion of Kraft Lignin into Liquid Fuels with a Co-Zn-Beta Zeolite Catalyst. *Appl. Catal. B* **2020**, *268*, 118429. [[CrossRef](#)]
28. Du, B.; Chen, C.; Sun, Y.; Yu, M.; Yang, M.; Wang, X.; Zhou, J. Catalytic Conversion of Lignin to Bio-Oil over PTA/MCM-41 Catalyst Assisted by Ultrasound Acoustic Cavitation. *Fuel Process. Technol.* **2020**, *206*, 106479. [[CrossRef](#)]
29. Nandiwale, K.Y.; Danby, A.M.; Ramanathan, A.; Chaudhari, R.v.; Subramaniam, B. Zirconium-Incorporated Mesoporous Silicates Show Remarkable Lignin Depolymerization Activity. *ACS Sustain. Chem. Eng.* **2017**, *5*, 7155–7164. [[CrossRef](#)]
30. Kim, J.-Y.; Park, S.Y.; Choi, I.-G.; Choi, J.W. Evaluation of Ru<sub>x</sub>Ni<sub>1-x</sub>/SBA-15 Catalysts for Depolymerization Features of Lignin Macromolecule into Monomeric Phenols. *Chem. Eng. J.* **2018**, *336*, 640–648. [[CrossRef](#)]
31. Limarta, S.O.; Kim, H.; Ha, J.-M.; Park, Y.-K.; Jae, J. High-Quality and Phenolic Monomer-Rich Bio-Oil Production from Lignin in Supercritical Ethanol over Synergistic Ru and Mg-Zr-Oxide Catalysts. *Chem. Eng. J.* **2020**, *396*, 125175. [[CrossRef](#)]
32. Tran, M.H.; Phan, D.-P.; Nguyen, T.H.; Kim, H.B.; Kim, J.; Park, E.D.; Lee, E.Y. Catalytic Hydrogenolysis of Alkali Lignin in Supercritical Ethanol over Copper Monometallic Catalyst Supported on a Chromium-Based Metal–Organic Framework for the Efficient Production of Aromatic Monomers. *Bioresour. Technol.* **2021**, *342*, 125941. [[CrossRef](#)]
33. Luo, B.; Zhou, L.; Tian, Z.; He, Y.; Shu, R. Hydrogenolysis of Cornstalk Lignin in Supercritical Ethanol over N-Doped Micro-Mesoporous Biochar Supported Ru Catalyst. *Fuel Process. Technol.* **2022**, *231*, 107218. [[CrossRef](#)]
34. Jiang, W.; Cao, J.-P.; Zhu, C.; Xie, J.-X.; Zhao, L.; Zhang, C.; Zhao, X.-Y.; Zhao, Y.-P.; Bai, H.-C. Selective Hydrogenolysis of C–O Bonds in Lignin and Its Model Compounds over a High-Performance Ru/AC Catalyst under Mild Conditions. *Chem. Eng. Sci.* **2022**, *253*, 117554. [[CrossRef](#)]
35. Su, S.; Xiao, L.; Chen, X.; Wang, S.; Chen, X.; Guo, Y.; Zhai, S. Lignin-First Depolymerization of Lignocellulose into Monophenols over Carbon Nanotube-Supported Ruthenium: Impact of Lignin Sources. *ChemSusChem* **2022**, *15*, e202200365. [[CrossRef](#)] [[PubMed](#)]
36. Guo, J.-P.; Liu, F.-J.; Bie, L.-L.; Si, X.-G.; Li, Y.-H.; Song, P.; Liu, N.; Zhao, Y.-P.; Huang, Z.-X.; Cao, J.-P.; et al. Selective Cleavage of C–O Bond in Lignin and Lignin Model Compounds over Iron/Nitrogen Co-Doped Carbon Supported Ni Catalyst. *Fuel* **2022**, *316*, 123338. [[CrossRef](#)]
37. Jiang, W.; Cao, J.-P.; Yang, Z.; Xie, J.-X.; Zhao, L.; Zhu, C.; Zhang, C.; Zhao, X.-Y.; Zhao, Y.-P.; Zhang, J.-L. Hydrodeoxygenation of Lignin and Its Model Compounds to Hydrocarbon Fuels over a Bifunctional Ga-Doped HZSM-5 Supported Metal Ru Catalyst. *Appl. Catal. A Gen.* **2022**, *633*, 118516. [[CrossRef](#)]
38. Wu, J.; Zhu, X.; Fu, Y.; Chang, J. Study on Selective Preparation of Phenolic Products from Lignin over Ru–Ni Bimetallic Catalysts Supported on Modified HY Zeolite. *Ind. Eng. Chem. Res.* **2022**, *61*, 3206–3217. [[CrossRef](#)]
39. Zeng, Z.; Xie, J.; Guo, Y.; Rao, R.; Chen, B.; Cheng, L.; Xie, Y.; Ouyang, X. Hydrogenolysis of Lignin to Produce Aromatic Monomers over Fe Pd Bimetallic Catalyst Supported on HZSM-5. *Fuel Process. Technol.* **2021**, *213*, 106713. [[CrossRef](#)]
40. Xu, Y.; Chen, P.; Lv, W.; Wang, C.; Ma, L.; Zhang, Q. Hydrogenolysis of Organosolv Hydrolyzed Lignin over High-Dispersion Ni/Al-SBA-15 Catalysts for Phenolic Monomers. *Chin. J. Chem. Eng.* **2021**, *32*, 307–314. [[CrossRef](#)]
41. Luo, B.; Huang, Z.; Shu, R.; Cheng, Y.; Tian, Z.; Wang, C.; Chen, Y. Effects of Metal Sites and Acid Sites on the Hydrogenolysis of Cornstalks in Supercritical Ethanol during Lignin-First Fractionation. *Sustain. Energy Fuels* **2021**, *5*, 6097–6106. [[CrossRef](#)]
42. Jiang, W.; Cao, J.-P.; Xie, J.-X.; Zhao, L.; Zhang, C.; Zhao, X.-Y.; Zhao, Y.-P.; Zhang, J.-L. Catalytic Hydrodeoxygenation of Lignin and Its Model Compounds to Hydrocarbon Fuels over a Metal/Acid Ru/HZSM-5 Catalyst. *Energy Fuels* **2021**, *35*, 19543–19552. [[CrossRef](#)]
43. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-L. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [[CrossRef](#)]
44. Zhao, C.; Huang, J.; Yang, L.; Yue, F.; Lu, F. Revealing Structural Differences between Alkaline and Kraft Lignins by HSQC NMR. *Ind. Eng. Chem. Res.* **2019**, *58*, 5707–5714. [[CrossRef](#)]
45. Lin, F.; Ma, Y.; Sun, Y.; Zhao, K.; Gao, T.; Zhu, Y. Heterogeneous Ni–Ru/H-ZSM-5 One-Pot Catalytic Conversion of Lignin into Monophenols. *Renew Energy* **2021**, *170*, 1070–1080. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.