Proceedings of the ASME 2019
Internal Combustion Engine Fall Technical Conference
ICEF2019
October 20-23, 2019, Chicago, IL, USA

ICEF2019-7170

Screening Compounds for Fast Pyrolysis and Catalytic Biofuel Upgrading Using Artificial Neural Networks

Travis Kessler[1], Thomas Schwartz[2], Hsi-Wu Wong[3], and J. Hunter Mack[4]

[1]Department of Electrical and Computer Engineering, University of Massachusetts Lowell
[2]Department of Chemical and Biomedical Engineering, University of Maine
[3]Department of Chemical Engineering, University of Massachusetts Lowell
[4]Department of Mechanical Engineering, University of Massachusetts Lowell

ABSTRACT

There is significant interest among researchers in finding economically sustainable alternatives to fossil-derived drop-in fuels and fuel additives. Fast pyrolysis, a method for converting biomass into fuels and fuel additives, is a promising technology with the potential to be two to three times less expensive at scale when compared to alternative approaches such as gasification and fermentation. Selective pyrolysis, via micro-mixing manipulation and chemical catalysis, allows resulting bioblendstocks to be tuned to perform optimally in diesel engines; however, many bio-oils derived from fast pyrolysis have a high oxygen content and high acidity, indicating poor performance in diesel engines when used as fuels or fuel additives. The variance in performance for derived compounds introduces a feedback loop in researching acceptable fuels and fuel additives, as various combustion properties for these compounds must be determined after pyrolysis and catalytic upgrading occurs. The present work aims to reduce this feedback loop by utilizing artificial neural networks to preemptively screen compounds that will be produced from fast pyrolysis and catalytic upgrading. Specifically, the cetane number and sooting propensity of derived compounds is predicted, and the viability of these compounds as fuels and fuel additives is analyzed.

1 INTRODUCTION

The emphasis on finding renewable and cleaner sources of energy has become prevalent throughout the greater scientific community due to concerns regarding global climate change and decreasing reserves of traditional petroleum-based fuels. Consequently, researchers have focused their efforts on discovering novel fuels derived from renewable sources such as biofuels derived from plant matter. One method for converting biomass into bio-oil is fast pyrolysis, however not all compounds derived from fast pyrolysis perform optimally in existing engines and/or produce high amounts of negative byproducts. This highlights the need to screen compounds that will be produced from fast pyrolysis before the procedure occurs. To reduce the inherent feedback loop associated with alternative fuel research, machine learning (specifically artificial neural networks, ANN's) can be employed to predict key properties of the products of fast pyrolysis.

1.1 Cetane Number

Cetane number (CN) measures a fuel's ignition quality in a diesel engine and is derived from the ignition delay after the fuel is injected. Diesel fuel typically has a CN of 40-50. Two common techniques for mea-

suring CN are with an Ignition Quality Tester (IQT) and a Cooperative Fuel Research (CFR) engine. An IQT ascertains the ignition delay of a given fuel in a constant volume combustion chamber by measuring the time between injection and combustion [1]. Alternatively, a CFR utilizes two reference compounds, n-hexadecane and isocetane (with CN's of 15 and 100, respectively) – igniting a given fuel in an equivalent blend of these compounds results in a volume fraction of these compounds linear related to the fuel's CN [2]. A CFR is preferred over the IQT, as the CFR reflects typical engine behavior; however, the IQT requires less test fuel (around 100 mL). While methods like the IQT utilize a small volume of fuel for determining CN, the number of tests required for determining CN's of a suite of potential alternative fuels results in a considerable time and monetary investment.

Using computational techniques to predict CN has a pervasive history and includes a variety of methods. Such methods include consensus modeling, where linear and non-linear models are employed in parallel to obtain an averaged predicted CN value – these models can predict CN for a variety of molecular classes with a blind prediction root-mean-squared error of 6.5 [3], however multiple linear models are surpassed by the accuracy of ANN's when fatty acid methyl esters are used to predict CN [4]. Many methods rely on quantitative structure-property relationship (QSPR) values, which are numerical measurements of an assortment of physical and chemical properties relating to molecules, and have proven to be successful when applied to predicting the CN of pure hydrocarbons [5] and branched paraffins [6].

In accordance with these findings, this paper utilizes feed-forward ANN's constructed using the backpropagation algorithm and trained with QSPR values. ANN's provide a non-linear model architecture, allowing a multidimensional input vector containing a suite of individual QSPR values to be correlated to an experimental property value. QSPR values are utilized due to the wide range of physical and chemical property representations available, subsequently distinguishing one molecule from another [7]. Additionally, it has recently been shown that ANN's can extend their predictive capabilities to a variety of molecular classes, including predicting the CN of biomass-derived furanic compounds [8].

1.2 Yield Sooting Index

Various sooting indices are used to measure soot formation and how much particulate matter is emitted by a fuel during combustion. The Threshold Sooting Index (TSI) was developed to standardize the measurement of soot through smoke point and ranks fuels on a 0-100 scale using reference molecules [9]. Measuring smoke point involves measuring the maximum flame height attainable by a fuel combusting in a test lamp without smoking and has been shown to be a dependable indicator of sooting propensity of aviation fuels in turbines [10] as well as emissions from SI engines [11]. To account for reduced stoichiometric air required by oxygenated fuels, the oxygen extended sooting index (OESI) was defined as an extension to TSI [12]. While these smaller, bench-scale methods of measuring sooting propensity through smoke point are convenient, they suffer from a few disadvantages; operator bias in estimating an appropriate flame shape may occur, as well as the requirement of up to 20 mL of the fuel in order to measure [10].

Due in part to these disadvantages the yield sooting index (YSI) was developed, whose measurement is not based on smoke point, rather the maximum soot volume fraction measured in a flame ignited by a fuel doped with the molecule of interest [12]. YSI is measured on a 0-100 scale, using reference fuels n-hexane and benzene with YSI values of 0 and 100 respectively. Diesel fuel typically has a YSI of 235-250. This method of measurement requires a significantly smaller volume of the sample, and correlates adequately with TSI [13] [14] [15] [16]. A "high sooting" YSI scale was also developed to emphasize mass-fraction-based fuel doping, and recent research has provided a unified YSI scale, standardizing measurements from a variety of compounds and compound groups [17].

Recently, ANN's have been applied to predicting YSI using QSPR values and obtained 95% confidence in blind prediction accuracy [18]. Due to these promising results, this paper applies the same model architecture it utilizes for CN predictions to predict YSI (feed-forward ANN's constructed with backpropagation in conjunction with QSPR values). The unified YSI scale is used, as the models constructed aim to predict YSI for a variety of compounds and compound groups.

1.3 Fast Pyrolysis

Fast pyrolysis converts biomass into bio-oils by rapidly heating (>1000 °C/sec) the biomass at temperatures ranging from 400-600 °C for between 1-5 seconds and may yield up to 70 wt% of liquid phase products (bio-oils) [19]. It has been shown in recent techno-economic analyses that the cost to produce bio-oils via fast pyrolysis is comparable to the cost of producing bio-oils via alternative methods such as gasification, indicating fast pyrolysis is viable at scale

[20]. The products of fast pyrolysis can range from alcohols and ethers to furanic and phenolic compounds depending on the feedstock utilized in the [21]. Some products that form as a result of using lignocellulosic biomass as the primary feedstock in fast pyrolysis have been identified as potential replacements for fossil-derived fuel additives and drop-in fuels [22], however many of these compounds have a high oxygen content (indicating poor performance as a drop-in fuel additive). The CN and YSI of these phenolic bio-oils are evaluated in the present work.

1.4 Catalytic Upgrading

Catalytic upgrading is significant because it increases the hydrogen-to-carbon ratio, reduces oxygen content and increases molecular weights for oxygenated compounds derived from fast pyrolysis through a process called hydrodeoxygenation [23]. Many hydrocarbon products were observed as a result of hydrotreating oxygen-rich compounds using Pd/C catalysts at 310-375 degrees Celsius and then subjecting them to hydrocracking at 400 degrees Celsius and 2000 psig [24]. Hydrogenolysis of phenolic compounds using Ru/TiO2 catalysts results in aromatic ring hydrogenation occurring on the Ru/Ti nanoparticle surface, in turn forming compounds with fewer oxygen atoms and that are aromatic in nature (both important qualities for drop-in fuel additives) [25] [26]. The present work focusses on evaluating the CN and YSI of compounds resulting from performing hydrogenolysis on phenolic bio-oil using Ru/TiO2 catalysts.

2 MATERIALS AND METHODS

2.1 Experimental Data

Experimental CN data was obtained from the NREL Compendium of Experimental Cetane Number data [27] and other sources [3] [28] [29] totaling in 445 unique compounds. Methods for obtaining experimental CN values include derivations from blend measurements, the use of an IQT/CFR and other, unknown ignition delay methods. Most experimental values were obtained using an IQT/CFR, as these methods are more accurate than blending/other methods. Experimental YSI data was obtained from a variety of sources [13] [14] [15] [16], was measured using laser-induced incandescence and other gas-phase steady-state measurement techniques, and totaled in 421 compounds.

Simple molecular-input line-entry system (SMILES) strings were obtained for all compounds comprising the CN and YSI databases using MarvinSketch [30] and validated using compound entries on PubChem [31]. SMILES strings were then converted to MDL Molfiles using Open Babel to generate three-dimensional geometry for the compound [32], and the Molfiles were fed into PaDEL-Descriptor to generate 1444 1D/2D and 431 3D QSPR descriptors for each compound [33]. QSPR values and experimental property values comprise input and target data respectively during ANN training.

Table 1 displays a list of compounds that are expected be produced by performing fast pyrolysis on lignocellulosic biomass, and includes phenolic compounds, furanic compounds and benzenes. Table 2 illustrates the expected products when hydrogenolysis using Ru/TiO2 catalysts is performed on each compound in Table 1 with the inclusion of phenol in the catalysis process. For each compound listed in Tables 1 and 2, QSPR descriptors were generated using the previously mentioned techniques.

Table 1. Products of fast pyrolysis using lignocellulosic feed stock

Compound ID	Compound Name
1	phenol
2	methanol
3	acetol
4	2-furanmethanol
5	5-hydroxy methyl furfural
6	2-hydroxy-cyclopent-2-en-1one
7	2-methoxy phenol
8	2-methyl phenol
9	4-methyl phenol
10	2-methoxy-4-methyl phenol
11	3,5-dimethyl phenol
12	3-ethyl phenol
13	4-ethyl-2-methoxy phenol

Table 2. Products of catalytic upgrading of Table 1 molecules with phenol

Table 1 ID	Upgrade with Phenol
1	Diphenyl Ether
2	Methoxybenzene
3	1-phenoxy-2-propanone
4	2-phenoxymethylfuran
5	5-phenoxymethylfurfural
6	2-phenoxy-2-cyclopentene-1-one
7	1-methoxy-2-phenoxybenzene
8	1-methyl-2-phenoxybenzene
9	1-methyl-4-phenoxybenzene
10	2-methoxy-4-methyl-1-phenoxybenzene
11	1,3-dimethyl-5-phenoxybenzene
12	1-ethyl-3-phenoxybenzene
13	1-ethyl-4-phenoxybenzene

2.2 QSPR Descriptor Selection

Random forest regression using Scikit-learn was employed to determine QSPR descriptor "importance", a numerical measurement indicating correlation between descriptor values and experimental property values for CN and YSI – the higher the importance of a descriptor, the more it contributes to a correlation to a given property [34]. The sum of all descriptor importances is equal to one. Figure 1 illustrates ANN performance (RMSE of CN) versus the number of important descriptors added as inputs to the ANN. Performance degrades significantly past 500 descriptor additions, and it was found that 15-25 descriptors balances computation time and predictive accuracy. Degrading performance can be attributed to the number of constant-value descriptors, and are pernicious to the ANN in that the ANN is unable to determine any relationship between them an a given experimental value. For example, "Nn", or the number of nitrogen atoms present in the compound, plays no significant role during training as all training data is comprised of hydrocarbons. Tables 3 and 4 show the selected QSPR descriptors for CN and YSI respectively.

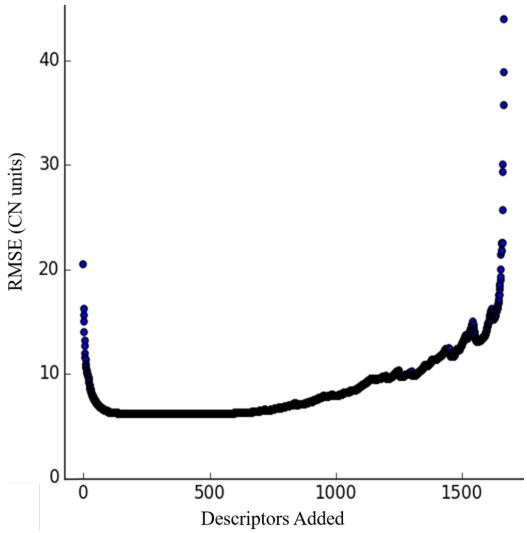

Figure 1. ANN performance versus number of important descriptors added to the ANN's input

Table 3. Descriptors and importances for CN

Name	Imp.	Description
RotBFrac	0.5747	Fraction of rotatable bonds, excluding terminal bonds
AVP-5	0.052	Average valence path, order 5
GATS2e	0.0159	Geary autocorrelation - lag 2 / weighted by Sanderson electronegativities
GATS2m	0.0137	Geary autocorrelation - lag 2 / weighted by mass
GATS2c	0.0121	Geary autocorrelation - lag 2 / weighted by charges
ATSC2c	0.0083	Centered Broto-Moreau autocorrelation - lag 2 / weighted by charges
AVP-6	0.0081	Average valence path, order 6
AlogP	0.0074	Ghose-Crippen LogKow
SssCH2	0.0066	Sum of atom-type E-State: -CH2-
ALogp2	0.0063	Square of ALogP
ATSC2m	0.0059	Centered Broto-Moreau autocorrelation - lag 2 / weighted by mass
ATSC1e	0.0051	Centered Broto-Moreau autocorrelation - lag 1 / weighted by Sanderson electronegativities
RDF140v	0.0044	Radial distribution function - 140 / weighted by relative van der Waals volumes

TDB3u	0.0041	3D topological distance based autocorrelation - lag 3 / unweighted
RDF140p	0.0041	Radial distribution function - 140 / weighted by relative polarizabilities

Table 4. Descriptors and importances for YSI

Name	Imp.	Description
piPC5	0.2602	Conventional bond order ID number of order 5
piPC4	0.1248	Conventional bond order ID number of order 4
piPC3	0.1099	Conventional bond order ID number of order 3
TpiPC	0.0975	Total conventional bond order (up to order 10)
piPC6	0.0853	Conventional bond order ID number of order 6
SpMax1_Bhm	0.0575	Largest absolute eigenvalue of Burden modified matrix - n 1 / weighted by relative mass
SpMax1_Bhv	0.0291	Largest absolute eigenvalue of Burden modified matrix - n 1 / weighted by relative van der Waals volumes
R_TpiPCTPC	0.0201	Ratio of total conventional bond order (up to order 10) with total path count (up to order 10)
SpMax1_Bhp	0.0195	Largest absolute eigenvalue of Burden modified matrix - n 1 / weighted by relative polarizabilities
ETA_Eta_F	0.0181	Functionality index EtaF
piPC8	0.0133	Conventional bond order ID number of order 8
piPC7	0.0128	Conventional bond order ID number of order 7
SpMin1_Bhi	0.0126	Smallest absolute eigenvalue of Burden modified matrix - n 1 / weighted by relative first ionization potential
MLFER_E	0.0121	Excessive molar refraction
ETA_Beta	0.0100	A measure of electronic features of the molecule

2.3 ANN Hyperparameter Tuning

ANN's were trained using the Adam optimization function, which possesses five hyperparameters that affect the quality of the ANN's training [35]:

- β_1: exponential decay rate 1 for moment estimates
- β_2: exponential decay rate 2 for moment estimates
- ϵ: number to prevent division by zero in the algorithm implementation
- α: learning rate (i.e. step size)
- learning rate decay: degradation of learning rate after each learning epoch

In addition to these five training variables, the optimal number of neurons in the ANN's hidden layer(s) must be determined. Given the number of hyperparameters that must be manually tuned under normal circumstances, an artificial bee colony (ABC) was utilized to algorithmically determine the optimal sets of hyperparameters for both CN and YSI models. ABC's mimic the foraging behavior of honeybees to search a multidimensional search space of tunable variables and have been shown to out-perform genetic algorithms and other particle swarm optimization algorithms in tuning various hyperparameters in ANN's [36]. The ABC was supplied with a fitness function to determine the ability of ANN's to predict values for unseen values, where a lower RMSE for a given set of hyperparameters was deemed better performing. The ABC was run for 20 search cycles with 50 employer bees for both CN and YSI datasets. Tables 5 and 6 show the tuned hyperparameters given the use of the Adam optimization function and two hidden layers.

Table 5. Tuned hyperparameters for CN model

Hyperparameter	Value
β_1	0.0482
β_2	0.8011
ϵ	0.2752
α	1.0
Decay	0.2710
Neurons in 1st hidden layer	35
Neurons in 2nd hidden layer	31

Table 6. Tuned hyperparameters for YSI model

Hyperparameter	Value
β_1	0.3241
β_2	0.6434
ϵ	0.9485
α	0.2622
Decay	0.0145
Neurons in 1st hidden layer	39
Neurons in 2nd hidden layer	10

2.4 Candidate ANN Training/Selection

ANN training was performed with ECNet, an open source machine learning toolkit created to predict fuel properties of potential next-generation fuels [37]. A model is considered a collection of ANN's whose predictions are averaged to obtain a final prediction, producing a more accurate prediction in a similar fashion to classifier ensembles [38]. Each ANN was chosen from a pool of ANN candidates, where the pool's goal was to optimally predict either CN or YSI. The ANN that was chosen from each pool achieved the lowest RMSE in predicting unseen data across all the pool's candidates. 5 pools, each with 75 candidate ANN's were trained for both CN and YSI.

Each candidate ANN was supplied with a random learning set and a random validation set, 70% and 20% of the total data respectively, while the remaining 10% of data remained constant for all candidates to measure their performance in predicting unseen data. The ANN's were trained using the backpropagation algorithm and the learning set, while the validation set measured the progress of the ANN's learning. Once performance stopped improving on the validation set, learning was terminated – this was done to prevent any overfitting on the learning set. Shuffling learning/validation sets allowed each ANN to learn from a different representation of data, and each ANN's predictions are unique (often predicting slightly higher or slightly lower than the known experimental value). Consequently, the final prediction of the model tends to be more accurate. Model performance was measured by determining RMSE and the r-squared correlation coefficient when predicting unseen data (the constant test set).

3 RESULTS AND DISCUSSION

Figures 2 and 3 show parity plots between predicted values and experimental values for CN and YSI respectively, illustrating the performance of the training (learning and validation) set and the blind test set. The center dashed lines illustrate a 1:1 parity between predicted values and experimental values, and the outer dashed lines show bounds imposed by the test set's RMSE. Test set RMSE and r-squared are significantly better for both CN and YSI, as candidate ANN's were selected based on their ability to predict unseen data.

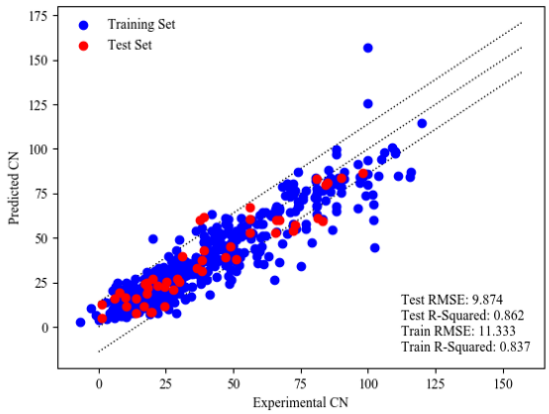

Figure 2. CN parity plot illustrating predicted values vs. experimental values

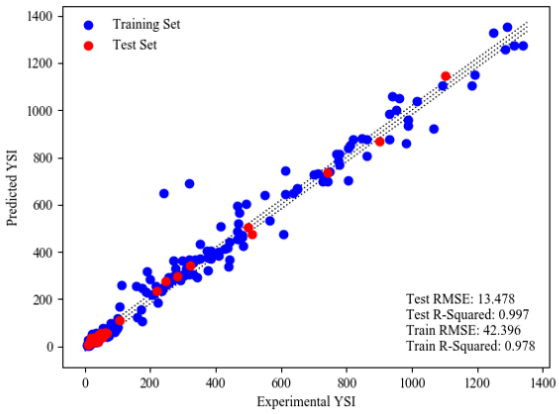

Figure 3. YSI parity plot illustrating predicted values vs. experimental values

Table 7 displays CN and YSI predictions for the products of fast pyrolysis outlined in Table 1, and Table 8 displays CN and YSI predictions for the products of catalytic upgrading outlined in Table 2. As seen by the disparity between pre- and post-catalytic upgrading CN values (Table 7 and Table 8), upgraded compounds such as diphenyl ether (a result of phenol/phenol upgrading), 1-methoxy-2-phenoxybenzene (a result of 2-methoxy phenol/phenol upgrading) and 2-methoxy-4-methyl-1-phenoxybenzene (a result of 2-methoxy-4-methyl phenol/phenol upgrading) have significantly higher CN's. Furan/phenol-based upgrades also have higher

CN's. Ethyl/phenol-based upgrades did not improve CN significantly. All upgraded compounds have significantly higher YSI's than their pre-upgraded products except for diphenyl ether.

Table 7. Predicted CN/YSI for products of fast pyrolysis

Name	CN	YSI
phenol	1.84	161.71
methanol	9.10	1.55
acetol	8.76	14.00
2-furanmethanol	7.16	81.89
5-hydroxy methyl furfural	7.02	215.34
2-hydroxy-cyclopent-2-en-1one	7.82	71.24
2-methoxy phenol	8.75	330.34
2-methyl phenol	2.44	237.00
4-methyl phenol	2.79	244.16
2-methoxy-4-methyl phenol	8.38	403.78
3,5-dimethyl phenol	1.42	357.99
3-ethyl phenol	5.60	315.51
4-ethyl-2-methoxy phenol	11.61	479.04

Table 8. Predicted CN/YSI for products of catalytic upgrading

Name	CN	YSI
Diphenyl Ether	12.01	168.85
Methoxybenzene	9.71	649.02
1-phenoxy-2-propanone	11.11	342.27
2-phenoxymethylfuran	12.11	451.96
5-phenoxymethylfurfural	12.94	525.03
2-phenoxy-2-cyclopentene-1-one	11.80	541.33
1-methoxy-2-phenoxybenzene	16.06	689.18
1-methyl-2-phenoxybenzene	6.77	684.55
1-methyl-4-phenoxybenzene	5.83	696.12
2-methoxy-4-methyl-1-phenoxybenzene	17.06	746.00
1,3-dimethyl-5-phenoxybenzene	5.75	747.66
1-ethyl-3-phenoxybenzene	8.90	714.75
1-ethyl-4-phenoxybenzene	8.51	718.21

4 CONCLUSION

From the predicted CN/YSI values displayed in Tables 7 and 8, it can be concluded that:

- Catalytically upgrading phenol, methoxy-phenolic and furanic compounds using Ru/TiO2 catalysts and phenol during the catalysis process yields compounds with higher CN values, likely attributed to the lower oxygen content of the products
- Catalytically upgrading ethyl-based compounds using Ru/TiO2 catalysts and phenol during the catalysis process yields compounds with no significant improvement in CN values
- Catalytically upgrading products of fast pyrolysis using Ru/TiO2 catalysts and phenol during the catalysis process yields compounds with significantly higher sooting propensity, with the exception of diphenyl ether

Based on these findings, further pursuit of catalytically upgrading phenolic/furanic compounds in an experimental setting is recommended. Additionally, different additives during the catalysis process besides phenol should be assessed, such as furanic additives.

Additional investigations into why the selected QSPR descriptors contribute to CN/YSI from a chemical standpoint, and what role upgraded compounds play in a mixture of traditional petroleum-based fuel should be performed.

References

[1] *ASTM D6890-16e1, Standard Test Method for Determination of Ignition Delay and Derived Cetane Number (DCN) of Diesel Fuel Oils by Combustion in a Constant Volume Chamber.* ASTM International, West Conshohocken, PA, 2016, www.astm.org

[2] *ASTM 613-16a, Standard Test Method for Cetane Number of Diesel Fuel Oil.* ASTM International, West Conshohocken, PA, 2016, www.astm.org

[3] D. A. Saldana, L. Starch, P. Mougin, B. Rousseau, L. Pidol, N. Jeuland, B. Creton. *Flash Point and Cetane Number Prediuctions for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods.* Energy and Fuels, Vol. 25, No. 9, pp. 3900-3908, 2011

[4] R. Piloto-Rodriguez, Y. Sanchez-Borroto, M. Lapuerta, L. Goyos-Perez, S. Verhelst. *Prediction of cetane number of biodiesel using artifical neural networks and multiple linear regression.* Energy Conversion and Management 65, pp. 255-261, 2013

[5] E. A. Smolenskii, V. M. Bavykin, A. N Ryzhov, O. L. Slovokhotova, I. V. Chuvaeva, A. L. Lapidus. *Cetane number of hydrocarbons: calculations using optimal tolological indices.* Russian Chemical Bulletin, Vol. 57, No. 3, pp. 461-467, 2008

[6] H. Yang, C. Fairbridge, Z. Ring. *Neural Network Prediction of Cetane Number ofr iso-Paraffins and Diesel Fuel.* Petroleum Science and Technology, Vol. 19, No, 5-6, pp. 573-586, 2001

[7] T. Sennott, C. Gotianun, R. Serres, M. Ziabasharhagh, J. H. Mack, R. W. Dibble. *Artifical neural network for predicting cetane number of biofuel candidates based on molecular structure.* ASME 2013 Internal Combustion Engine Division Fall Technical Conference, 2013

[8] T. Kessler, E. R. Sacia, A. T. Bell, J. H. Mack. *Predicting the Cetane Number of Furanic Biofuel Candidates Using an Improved Artificial Neural Network Based on Molecular Structure.* ASME 2016 Internal Combustion Engine Division Fall Technical Conference, 2016

[9] H. Calcote, D. Manos. *Effect of molecular structure on incipient soot formation.* Combust. Flame, vol. 49, pp. 289-304, 1983

[10] *D1322, A. Test Method for Smoke Point of Kerosine and Aviation Turbine Fuel.* 2015

[11] E. J. Barrientos, J. E. Anderson, M. M. Maricq, A. L. Boehman. *Particulate matter indices using fuel smoke point for vehicle emissions with gasoline, ethanol blends, and butanol blends.* Combust. Flame, vol. 167, pp. 308-319, 2016

[12] E. J. Barrientos, M. Lapuerta, A. L. Boehman. *Group additivity in soot formation for the example of C-5 oxygenated hydrocarbon fuels.* Combust. Flame, vol. 160, 2013

[13] C. S. McEnally, L. D. Pfefferle. *Improved sotting tendency measurements for aromatic hydrocarbons and their implications for naphthalene formation pathways.* Combust. Flame, vol. 148, pp. 210-222, 2007

[14] C. S. McEnally, L. D. Pfefferle. *Sooting tendencies of nonvolatile aromatic hydrocarbons.* P. Combust. Inst., vol. 32, pp. 673-679, 2009

[15] C. S. McEnally, L. D. Pfefferle. *Sooting Tendencies of Oxygenated Hydrocarbons in Laboratory-Scale Flames.* Environ. Sci. Technol., vol. 45, pp. 2498-2503, 2011

[16] D. D. Das, C. S. McEnally, L. D. Pfefferle. *Sooting tendencies of unsaturated esters in non-premixed flames* Combust. Flame, vol. 162, pp. 1489-1497, 2015

[17] D. D. Das, P. C. St. John, C. S. McEnally, S. Kim, L. D. Pfefferle. *Measurements and Prediction of Sooting Tendencies of Pure Hydrocarbons.* 2017

[18] P. C. St. John, P. M. Kairys, D. D. Das, C. S. McEnally, L. D. Pfefferle, D. J. Robichaud, M. R. Nimlos. *A quantitative model for the prediction of sooting tendency from molecular structure.* Energy and Fuels, 2017

[19] D. M. Alonso, J. Q. Bond, J. A. Dumesic. *Catalytic conversion of biomass to biofuels.* Green Chemistry, vol. 12, pp. 1493-1513, 2010

[20] R. P. Anex, A. Aden, F. K. Kazi, J. Fortman, R. M. Swanson, M. M. Wright, J. A. Satrio, R. C. Brown, D. E. Daugaard, A. Platon, G. Kothandaraman, D. D. Hsu, A. Dutta. *Techno-economic comparison of biomass-to-transportation fuels via pyrolysis, gassification, and biochemical pathways.* Fuel, vol. 89(S1), pp. S29-S35, 2010

[21] P. R. Patwardhan, J. A. Satrio, R. C. Brown, B. H. Shanks. *Influence of inorganic salts on the primary pyrolysis products of cellulose.* Bioresource Technology, vol. 101(12), pp. 4646-4655, 2010

[22] S. P. S. Chundawat, G. T. Beckham, M. E. Himmel, B. E. Dale. *Deconstruction of lignocellulosic biomass into fuels and chemicals*. Annual Review of Chemical and Biomolecular Engineering, vol. 2, pp. 121-145, 2011

[23] J. D. Adjaye, N. N. Bakhshi. *Production of hydrocarbons by catalytic upgrading of a fast pyrolysis bio-oil* Fuel Processing Technology, vol. 45(3), pp. 161-202, 1995

[24] D. C. Elliott, T.R. Hart, G. G. Neuenshwander, L. J. Rotness, A. H. Zacker. *Catalytic hydroprocessing of biomass fast pyrolysis bio-oil to produce hydrocarbon products*. Env. Prog. & Sustainable Energy, vol. 28(3), pp. 441-449, 2009

[25] C. Newman, X. Zhou, B. Goundie, I. T. Ghampson, R. A. Pollack, Z. Ross, M. C. Wheeler, R. W. Meulenberg, R. N. Austin, B. G. Frederick. *Effects on support identity and metal dispersion in supported ruthenium hydrodeoxygenation catalysts*. Applied Catalysts A: General, vol. 477, pp. 64-74, 2014

[26] R. C. Nelson, B. Baek, P. Ruiz, B. Goundie, A. Brooks, M. C. Wheeler, B. G. Frederick, L. C. Grabow, R. N. Austin. *Experiemntal and Theoretical Insights into the Hydrogen-Efficient Direct Hydrodeoxygenation Mechanism of Thenol over Ru/TiO2*. ACS Catalysts, vol. 5, pp. 6509-6523, 2015

[27] J. Yanowitz, M. A. Ratcliff, R. L. McCormick, J. D. Taylor, M. J. Murphy. *Compendium of Experimental Cetane Numbers*. NREL/TP-5400-61693, 2014

[28] J. Taylor, R. McCormick, W. Clark. *Report on the relationship between molecular structure and compression ignition fuels*. NREL Technical Report, 2014

[29] M. Dahmen, W. Marquardt. *A Novel Group Contribution Method for the Prediction of the Derived Cetane Number of Oxygenated Hydrocarbons*. Energy and Fuels, vol. 29(9), pp. 5781-5801, 2015

[30] ChemAxon, 2015. *MarvinSketch, Version 15.10.19.0*. http://www.chemaxon.com

[31] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton. *Pubchem 2019 update: improved access to chemical data*. Nucleic Acids Res. 2019 Jan 8; 47(D1):D1102-1109

[32] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchinson. *Open Babel: An opem chemical toolbox*. Journal of Cheminformatics, vol. 3, pp. 33, 2011

[33] Yap CW. *PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints*. Journal of Computational Chemistry. 32 (7): 1466-1474, 2011

[34] L. Breiman. *Random forests*. Journal of Machine Learning, vol. 45, pp. 5-32, 2001

[35] D. P. Kingma, J. Ba. *Adam: A Method for Stochastic Optimization*. International Conference for Learning Representations, 2015

[36] Z. G. Cam, S. Cimen, T. Yildirim. *Learning parameter optimization of Multi-Layer Perceptron using Artificial Bee Colony, Genetic Algorithm and Particle Swarm Optimization*. IEEE 13th International Symposium on Applied Machine Intelligence and Informatics, 2015

[37] T. Kessler, J. H. Mack. *ECNet: Large scale machine learning projects for fuel property prediction*. Journal of Open Source Software, 2(17), 401, 2017

[38] L. Kincheva, C. Whitaker. *Measures of diversity in classifier ensembles*. Machine Learning, vol. 51, pp. 181-207, 2003