Artificial intelligence-enabled fuel design

Artificial Intelligence (AI) Enabled Fuel Design

Kiran K. Yalamanchi, Andre Nicolle, S. Mani Sarathy

Abstract

Fuel composition plays an important role both in efficiency and effectiveness of engines. Combined with the engine variables, fuel can span a wide range of composition space, which makes it demanding to find an optimal composition. Artificial Intelligence (AI) algorithms are attracting significant interest for predicting complex phenomenon. In this chapter, a discussion is presented on exploiting the advantages presented by machine learning algorithms for fuel formulation. The present fuel modelling scenario and a holistic approach necessary for fuel optimization is first presented. A wealth of AI algorithms are available to make use of in fuel formulation. These algorithms are discussed in line with their application to fuel formulation and the literature of the explored space in this area is presented. Additionally, a discussion is presented on how AI also helps in assisting the traditional computational fluid dynamic and chemical kinetic analysis for an elaborate study of fuels. Fuel development is just a step in the entire engine innovation cycle, and a perspective of how the AI fits in to this scenario is presented.

1. Transportation fuels

1.1 Fuel representation

Transportation fuels are complex mixtures resulting from the blending of several streams involved in separation, conversion, upgrading and blending processes. Many properties of fuels are regulated to ensure a safe and adequate behavior during storage, blending, distribution and end-use. These properties range from fundamental macroscopic physico-chemical data (density, viscosity) which can be readily predicted using first-principle physical models to more complex engineering indicators such as flash point (FP) or octane number (ON) for which AI techniques may outperform traditional bottom-up physical modeling [1] at the expense of traditional chemical insight [2].

For most practical fuels, detailed chemical analysis is unavailable and only its PIONA (paraffinic, iso-paraffinic, olefinic, naphthenic, aromatic) composition is provided. Fuel formulation methodologies based on detailed composition, therefore, require a molecular reconstruction step. Deep learning (DL) approaches were recently shown to outperform classical fuel reconstruction method such as entropy maximization [3].

A more global fuel representation would lie in the formulation of physico-chemical surrogates, i.e., mixtures of a few pure species (typically less than 3 per chemical family) emulating real fuel properties [4] to enable fundamental well-controlled experiments on simplified multicomponent mixtures. A recent attempt to design a jet fuel surrogate through MLR and SVM models trained on semi-quantitative [5] reactive force field simulations has been recently reported [6].

Various molecular 1-3D representations such as SMILES, InChI or connectivity matrices allow to encode species information into an AI computational framework. From this information, molecular descriptors may be generated and linked to macroscopic properties through quantitative structure-property relationships (QSPR). This chapter provides an overview of various AI-based methods than can be used to predict fuel properties and thereby enable fuel design.

1.2 Fuel formulation workflow

Data science, which is considered as the fourth pillar of scientific discovery, is generally not part of the training of fuel scientists and dedicated ecosystems fusing computational modelling, virtual high-throughput screening and big data analytics are still under development [7]. Ideally, fuel formulation would base on an integrated multi-scale optimization [8] of fuel properties and production processes to minimize well-to-wheel fuel impact on environment while ensuring economic feasibility (Figure 1). This can be performed through the coupling of Computer-Aided Molecular Design (CAMD) and IC Engine simulations [9,10] or joint fuel synthetic path – single cylinder combustion optimization [11]. In this coupled workflow, fuel properties are typically evaluated by Group Contribution (GC) or Quantitative Structure-Property Relationships (QSPR) models [12] involving molecular descriptors. From validated QSPR models, an inverse-QSPR (i-QSPR) problem can be solved to identify species possessing a targeted property [13]. Advances in computer science have enabled virtual high throughput screening in which large databases are reduced to a small set of promising molecules for experimentalists to work on [14]. AI developments support this subsequent step through the automation of experimental measurements of fuel properties using robotics [15,16] which may minimize measurement variability [17].

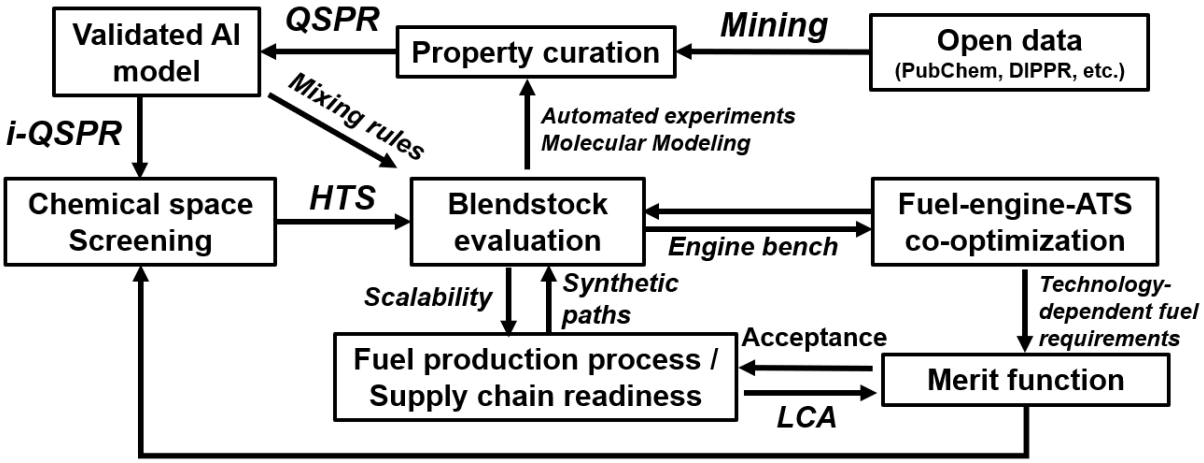


Figure 1 – Schematic overview of fuel formulation workflow from transportation fuel screening projects [18–22].

To avoid unbalanced fuel properties at later stages of the computational funnel [23] promising candidates for fuel/engine/aftertreatment (ATS) co-optimization can be selected by introducing a merit function assessing the combined effect of multiple fuel properties p_i on an overall

engine-dependent target, which can be engine brake efficiency η_b [21] according to the following expansion:

$$\frac{\Delta\eta_b}{\eta_b} = f(p_1, \ldots, p_n) = \sum_{i=1}^{n} \frac{1}{\eta_b} \frac{\partial \eta_b}{\partial p_i} \left(p_i - p_{i,ref}\right) \quad (1)$$

which highlights a coefficient $\frac{1}{\eta_b}\frac{\partial \eta_b}{\partial p_i}$ for each property. Co-optima project merit function involved RON, octane sensitivity, heat of vaporization, particulate index and catalyst light-off temperature. Merit function f is then maximized over the chemical space under molecular structure feasibility and properties constraints [24] using QSPR models previously developed for single or multiple fuel properties. It was found that RON exhibits the largest impact on the merit function of all the fuel properties considered, in line with previous findings.

To optimize fuel formulation more holistically, fuel well-to-wheel exergy [25–27]; or environmental impact factor may be used as alternate targets. An adaptive Neuro-Fuzzy Inference Systems (ANFIS) approach, a combination of neural networks and fuzzy systems, has been recently applied to biofuel life cycle analysis (LCA) to predict environmental index [28].

1.3 AI modeling approaches

Machine learning techniques used in fuel formulation involve mostly:

- Ensemble methods such as random forests (RF)
- Unsupervised techniques such as Principal Component Analysis (PCA) or k-means clustering
- Supervised techniques such as multiple linear regression (MLR) and support vector machines (SVM)
- Deep learning techniques such as convolutional neural (CNN) and recurrent neural (RNN) networks.
- Reinforcement learning (RL)

RF mimics human way of thinking by averaging predictions of several decision trees trained on different data samples using a greedy algorithm [29]. This ensemble average (wisdom of crowd effect) is also applied in consensus modeling with sometimes an additional optimization of weighting factors for the different models [30]. K-means clustering method, which groups similar data into evenly sized clusters, is seldom used as a stand-alone ML method for QSPR but is particularly useful during dataset preparation [31,32]. PCA is a method of choice for exploratory analysis as it performs dimension reduction by regressing the input data along the eigenvectors associated with the largest eigenvalues of the covariance matrix. However, it has difficulty accounting for non-linear relationships between molecular descriptors and chemical and technological properties. Furthermore, it cannot properly consider data that have strong non-orthogonal features or with high variance [33].

MLR assumes a multiple linear relationship between the input variables and dependent output variable. Each of the input variables is multiplied by a slope term with an additional intercept term. The training set is used to determine the weights of the algorithm which minimize the

error between the plane corresponding to relationship and the data points. The validation set is used to tune the hyper-parameters such the algorithm structure and regularization parameters. The test set is then used to report the (mean absolute or root mean squared) error of the model on unseen data. Depending on available computational resources, this workflow can be extended to k-fold cross-validation workflow in which the entire dataset is divided into k sets, each set being used as validation and test set once in a loop. MLR fails to capture non-linear behavior (e.g. azeotrope) but it is advisable to start modeling with MLR before developing a more complex model, unless there is an obvious reason not to do so. Linear models such as MLR are expected to be restricted to the description of the influence of homogeneous descriptors on similar compounds properties over a narrow variation intervals in the property space [34].

Fortunately, non-linearity is readily handled by SVM which designs the best hyperplane separating data into classes by maximizing its distance with the closest points (called support vectors). The performance metrics for SVM classification model are derived from the confusion matrix (true / false positive / negative rates). SVM employs custom kernel functions to map the labeled points of the training set in a higher dimension space where they can be linearly separated. This concept has been applied to kernel partial least squares (KPLS) regression method, which has been implemented industrially [35].

While SVM segregates data in sectors separated by the maximum distance, implying a limited precision in continuous quantities representation, artificial neural networks (ANN) map continuously input data into outputs. Even if their respective performances are problem dependent [36] and they do not provide any physical intuition, ANN are less sensitive to noisy data than SVM and allow a straightforward merit function implementation [37]. Due to their ability to adapt to highly non-linear relations, deep learning techniques are increasingly used in chemistry [38], although they are expected to require more data than classical ML methods for their training for the same fuel property [39].

A neural network consists of an input layer, several hidden layers of computational nodes and an output layer, corresponding to function composition [40]. Each of the nodes in one layer is interconnected with each and every other node of the next layer as shown in Figure 2. The nodes are created from the original inputs by a weighted sum, and then passed through an activation function f (rectified linear unit, sigmoid, tanh, etc) that allows for non-linearity in ANN. The training of the network involves finding the summation weights w_{ij} for the nodes in all layers. This is done by backpropagating the loss function gradient with respect to the weights using the chain rule from the fast to the first layer [41]. The tuple of hyper-parameters (number of hidden layers, number of nodes in each of the hidden layer, activation function, regularization parameter, learning rate) needs to be carefully chosen while training, either by intuition or dedicated (e.g. Bayesian or evolutionary) optimization algorithm.

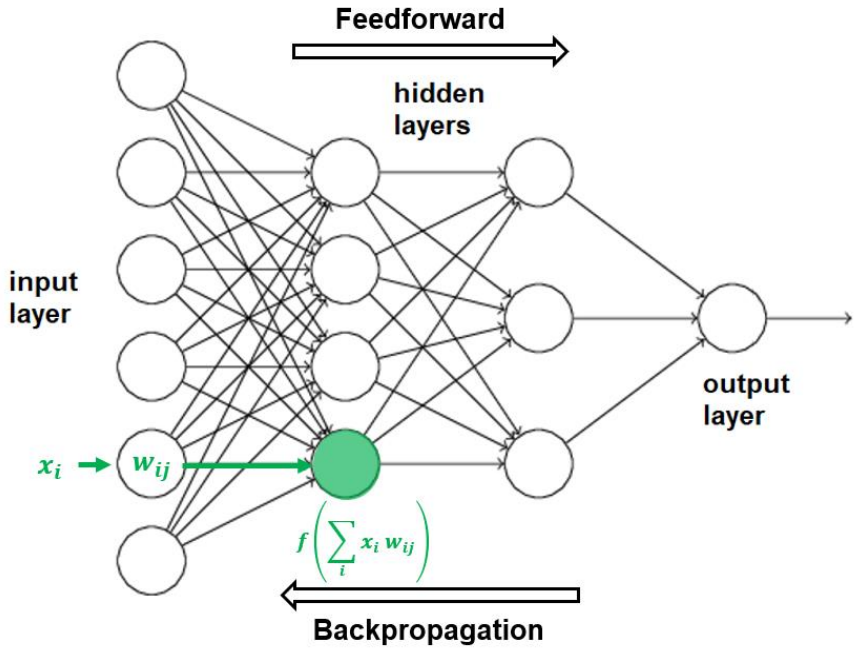


Figure 2 - Structure of artificial neural network (ANN).

Convolutional neural networks (CNN) mimic visual cortex connectivity by involving successively convolutional, pooling and perceptron layers which respectively filter, reduce feature dimensionality by cluster output combination and classify the image obtained. Recurrent Neural networks (RNN) were developed to model brain processes like memory and attention. RNN inputs have no determined size limit, allowing temporal dynamic behavior. RNN learns typically to predict the probability of the next character in a current SMILES based on previous ones [42]. Particular cases of RNN are long short-term memory (LSTM) networks which have an additional forget gate and gated recurrent units (GRU) which have a reset and an update gate. Blends may be handled through dedicated permutation-invariant network architecture including mix function to the interact / aggregate blocks and allowing to learn simultaneously on feature description and composition [43].

At the end of the day, the choice of a modeling approach will depend on the amount of data and the expected complexity of the input-output relationship. Usually, simple (but not too simple) models are preferable if the dataset is small. On the other hand, complex models perform generally better on larger datasets at the expense of large computational training cost and more propensity to overfitting, i.e. tendency to predict too closely a particular set of data at hand which results in unreliable prediction of new observations. This can be rationalized by monitoring both training R^2 (which approaches unity with increases model complexity) and validation Q^2 (which typically reaches a maximum with model complexity) squared correlation coefficients.

2. Application of AI to fuel formulation

2.1 High Throughput Screening: finding a needle in the haystack

Chemical space corresponds to the space spanned by any atom combination, estimated to be of the order of 10^{60} for limited weight CHONS compounds [44]. However, ML models are expected to work correctly only within the boundaries of their training dataset, calling for large

and more diverse datasets. Experiments or molecular modeling may be performed in the first place to augment this training dataset using a design of experiments (DoE) method [45,46]. Despite the much lower computational cost of ML models for prediction than training, chemical space enumeration is currently limited to 15-20 heavy atoms due to the combinatorial explosion [47]. Various alternate chemical screening methods were proposed such as molecular morphing [48] which generates structures connecting the starting and target structure.

To avoid exhaustive enumeration, various mutations / crossover may be applied to initial structure to produce new candidate structures [49], notably through SMARTS formalism, and evaluate the fitness of each molecule in the population [50]. The mutations on a structure are stopped if the said structure exceeds a threshold molecular weight or carbon number, similar to Lipinski and Morowitz rules featured in other fields of chemical research [51]. Figure 3 shows one such threshold on the CHO chemical space explored in transportation fuels components screening on Van Krevelen diagram [52].

Traditional chemical space screening proceeds through topological descriptors and use a molecular graph generator [53]. However, descriptor generation is nondifferentiable, preventing inverse mapping from the descriptor space back to molecules [54]. In deep learning approaches, molecules are typically generated from most probable latent space decoding [55], the differentiability of latent representations allowing gradient-based optimization. While RNN can generate SMILES through decoder, generative adversarial networks (GAN), combining a generator and a discriminator network, can directly generate molecular graphs. RNN-based RL allows to perform genuine QSPR analysis [56], one part of the network generating SMILES and the other predicting properties. Unlike RNN, GAN exempt the user from knowing probability distribution explicitly which is interesting for high-dimensional data [42].

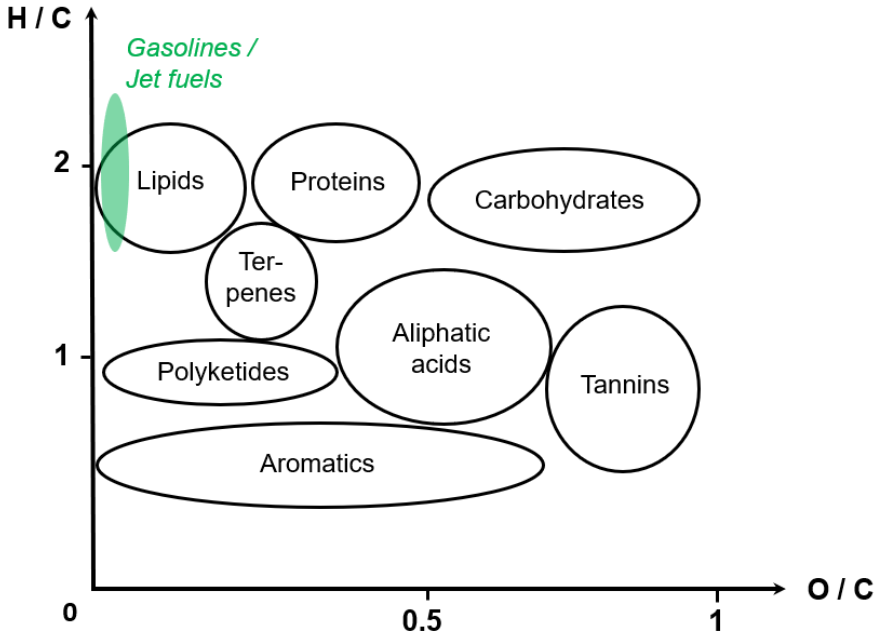


Figure 3 - Van Krevelen diagram representing schematically the CHO chemical space explored in transportation fuels components screening.

2.2 Fuel property prediction by ML models

Table 1 presents different approaches attempted to predict fuel properties, highlighting some molecular descriptors common to multiple properties. Most of the studies listed focus on pure species, mixture properties being deduced in turn through a mixing rule at the molecular descriptor or property level [57,58]. Also, few studies focus on simultaneous optimization of multiple fuel end-use properties [59].

Table 1 – Main approaches attempted to predict selected fuel properties and associated descriptors.

Species property (P_i)	AI approach	Reported descriptors [1]
Octane / Cetane Number, autoignition metrics	ANN based group contributions [60], SVM based on Boruta features elimination [61], CNN [62,63], Graph NN [64], ANN [65], k-NN, RF [66], HDMR/CNN [67,68]	Molecular weight, critical volume, Balaban / Kier-Hall / Wiener index, water/octanol partitioning.
Phase equilibria : Vapor pressure, boiling point, critical properties	MLR [69], CNN [70], MLR, ANN [71], ANN [72,73], LSTM [74], Matrix completion method [75]	Kappa shape index, Wiener and Harary indices.
Rheology : density, viscosity, surface tension	Consensus SVM / ANN prediction [76], ANN metamodel of phenomenological model parameters [77]	Molecular weight, Randic index, Solvent-accesible surface area
Storage / material compatibility : Induction period, swelling	PCA-ANN [78], ANN-SVM-MLR [79]	Excitation-emission matrix, number of sp^3 C, number of C=C bonds
HHV, LHV, HoV	GC [80], SVM, ANN[30,81], Ant colony - PLS – MLR [82], MLR [83], GA-SVM [84,85]	Van der Waals surface area, number of carbons,
Soot index, exhaust aftertreatment activity	Bayesian inference of GC [86], PCA-ANN [87], SVM, RF, PLS [88]	LUMO – HOMO energy, functional groups
Flammability limits, flash point	SVM, ANN [58], SVM [89], ANN – particle swarm [90], kNN, RF, MLR, SVM [91]	Randic index, polarizability, burden matrix, number of sp^3 C
Toxicity	SVM [92],MLR, RNN [93], SVM, RF, gradient boosting, NB,linear discriminant analysis, CNN [94]	Wiener index, LUMO – HOMO energy

However, in absence of detailed composition, fuel properties may be correlated to readily available observables [95]. Mendes et. al. [96] used ASTM distillation curves as input features. Fourier transform infrared (FTIR) [97–100], flame emission [101], nuclear magnetic resonance (NMR) [102–104], dispersive fiber-optic Raman [105] and dielectric spectroscopy [106] as well as GCxGC [107] were processed to generate discriminative input features by different groups. The spectra are either directly or refined to construct the functional groups which are then used as input features. The later requires a fundamental knowledge on the spectra corresponding to certain functional groups and also selection of important functional groups

that affect fuel property. An example of spectra conversion to functional groups is shown Figure 4.

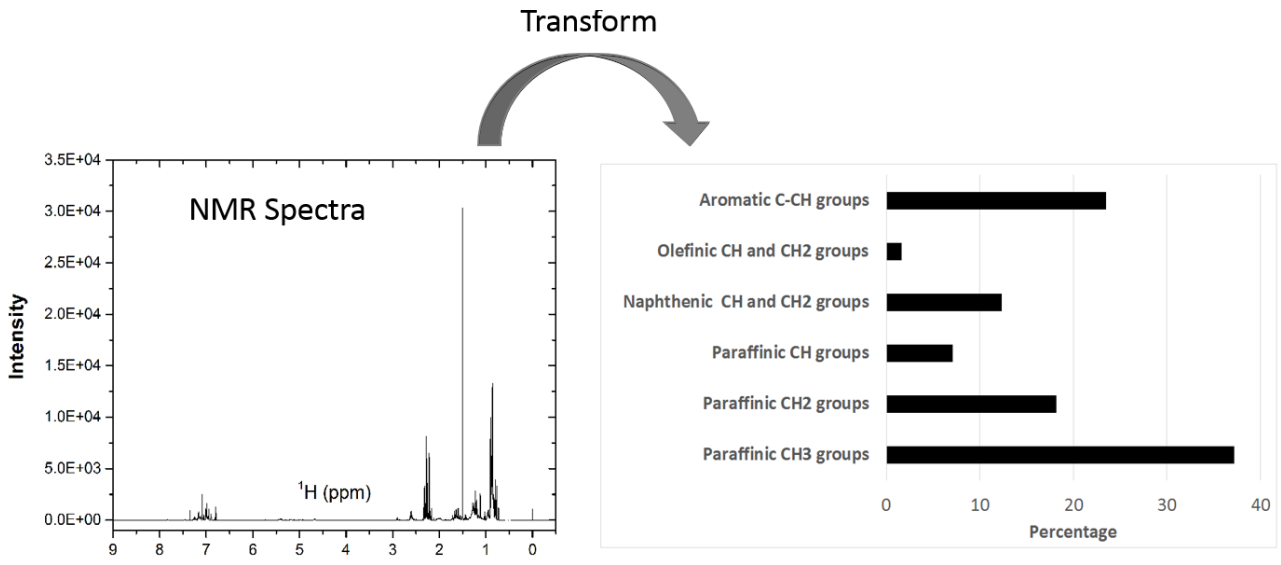
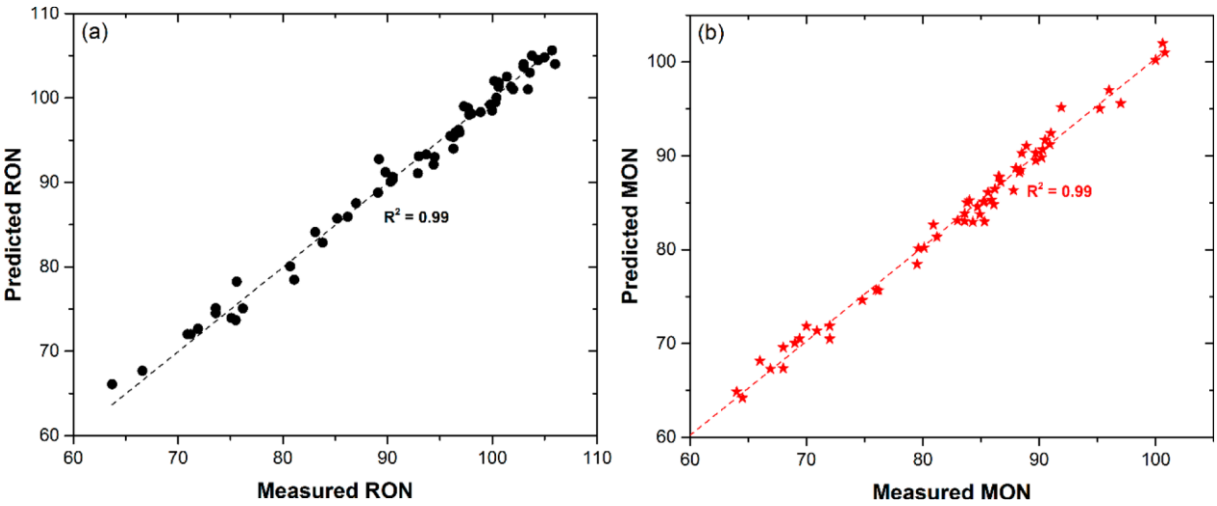


Figure 4- A sample of NMR spectra and its translation to functional groups.

Jameel et al. [104] generated ^{1}H NMR spectra of 128 pure hydrocarbons, 123 hydrocarbon ethanol blends of known composition, and 30 FACE (fuels for advanced combustion engines) gasoline ethanol blends. This was then converted to generate nine structural descriptors (paraffinic primary to tertiary carbons, olefinic, naphthenic, aromatic and ethanolic OH groups, molecular weight and branching index). ANN was then used to train RON and MON of the fuels in dataset with the generated structural descriptors. The hyper-parameters of ANN that were tuned in this study were the number of units per layer, regularization coefficients, and the number of layers, using the K-fold cross validation. Interestingly, while ANN modeling of RON and MON of these gasolines was successful (R^2=0.99 for both RON and MON), MLR training over the same dataset with same input features resulted in a poor correlation with R^2=0.51 for RON and R^2=0.52 for MON, as shown in Figure 5. This comparison demonstrates the non-linearity of fuel properties with respect to these structural descriptors requires a good model that can capture the non-linearity.



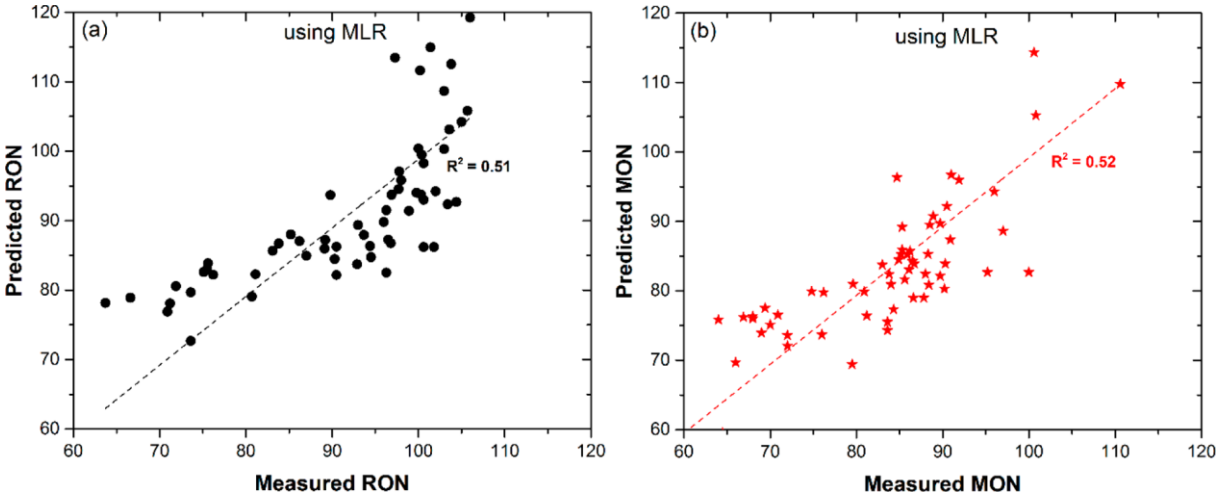


Figure 5- Comparison of measured and predicted (a) RON and (b) MON values for ANN and MLR models. This figure is reproduced from [104] with permission from ACS.

2.3 Reaction discovery

Designing a new fuel lies not only in finding structures exhibiting required physical property but it consists also in the screening of fuel synthesis, oxidation and pyrolysis chemical pathways to ensure that reactivity is controlled not only over a few regulatory conditions (e.g. RON, MON, induction delay, Flash point) but over the whole fuel life cycle and vehicle operation map. This step of the workflow allows an early identification of synergetic effects in fuel combustion and provides adequate combustion thermokinetic models for engine numerical design using computational fluid dynamic (CFD) and fuel property accuracy refinement.

Reaction network generation for fuel combustion or retrosynthesis may be performed by a rule-based expert system exploring a predefined template of reaction families for which the corresponding estimated rate constants may be estimated by on-the-fly quantum electronic structure calculations [108]. The search of optimal retrosynthetic paths may be performed through reinforcement learning [109]. Transition State geometries may be estimated from GC method [110] or Graph Neural Network [111] prior to Schrödinger equation solution, debottlenecking the $3N_{atom}$-6 dimensional potential energy surface (PES) screening and therefore kinetic rate constant generation. The electronic structure calculation may be made even faster by using a deep learning model to evaluate directly wavefunction [112]. It is also possible to train neural nets on post Hartree-Fock PES to achieve chemical accuracy, i.e error on electronic energy < 1 kcal/mol [113]. These models would be also useful during the abovementioned HTS step as it allows to generate quantum descriptors such as frontier orbitals energies at reduced cost [114]. To explore reaction pathways in a physics-agnostic way, quantum or classical Molecular Dynamics (MD) calculations may be performed to find unexpected products from starting molecule without defining reaction coordinate [115,116]. New Neural Network reactive force fields have been shown to be much more accurate than their conventional counterparts [117]. Further, a gaussian process regression of the atomistic PES model would allow a unified treatment for molecules and condensed-phase structures which would pave the way for a smoother exploration of multiphase engine pollutants and wall deposits evolution [118].

In any case AI-based methods are not expected to supersede physical methods which seek to explicitly understand the inner workings of a phenomenon, while data-driven methods focus on approximating its outer behavior [119].

2.4 Fuel-engine co-optimization

In a multiscale modeling framework, thermokinetic, transport and rheological information related to a new fuel derived from AI-enhanced physical models [120] may then be used in reactive CFD models. These fluid dynamics models use increasingly deep learning to estimate more accurately turbulent sub-grid scale fuel combustion rate than their algebraic counterparts [121,122].

Further, the computational burden associated with the stiff chemical kinetic source term integration in Navier Stokes equations may be largely reduced by either ANN-based HyChem kinetics [123] or direct deep learning on species profiles [124], resulting in typically 2 orders of magnitude speedup with respect to stiff chemical kinetics solver. An alternate approach training a CNN on CFD flow field data has been demonstrated recently, albeit on a single fuel [125].

Exhaust aftertreatment system needs also to be optimized given a new {fuel-engine} association. RF models have been successfully trained on heterogeneous kinetic Monte Carlo models for CO catalytic oxidation [126] and implemented in CatalyticFoam CFD code. An ANN model of a selective reduction catalyst coated on a particulate filter (SCRF) was trained on NEDC simulations using Axisuite aftertreatment simulation software, resulting in accurate predictions with a speedup of 450 [127]. However, the chemical space to screen to optimize catalytic formulation is much larger than for fuels due to a large variety of elements, polymorphs and crystal defects [128].

3. Conclusions and perspectives

As can be deduced from this short review of an expanding field, fuel discovery and formulation processes are increasing supported by AI-based methods, whether they touch upon data mining, robotics, chemical space exploration, integrated well-to-wheel fuel optimization, etc. While AI provides expanding possibilities to deal with complex physio-chemical phenomena involved during fuel life, numerous challenges remain to be met to make AI a daily tool for fuel experts.

First, there is currently no “Swiss army knife” which could address all problems involved in fuel formulation, since datasets for different properties are widely varying in size, homogeneity and linearity. Fuel experts have been typically juggling between different AI tools and datasets to cover the daunting range of scales, technological readiness levels and industrial constrains. Academic AI tools are progressively tested in industrial contexts, generating know-how for both communities, bridging the complexity and representativity gap.

Secondly, despite the intense research on fuel description, the current descriptors for mixtures are not fully satisfactory [22]. Similarly in the case of nanomaterials, the development of dedicated descriptors is clearly needed [129]. Yet, the unequivocal definition of mixtures is not a new problem in chemistry [130].

Thirdly, the evolution of the respective role of human researchers and AI in scientific discovery has to be addressed urgently to provide an engaging teleological view of researcher’s work.

Alan Turing noted that in ML the “teacher will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil's behavior” [131]. Despite progress in reinforcement learning [132] human intuition remains necessary to generate intelligibility out of ML model results and steer open-ended research [133]. Machines are certainly freeing up time for humans to solve higher-level questions but at the expense of redefining their connection with respect to the object of their research [134].

A branch in combustion that has not yet adopted AI is robotics. While control systems are largely used in combustion experiments, the experiments are planned and conducted with human interference. AI can be used to optimize the formulations of a certain mixture from the scratch using robotic experiments driven by machine learning [135]. These methods has been of interest lately for organic synthesis [136,137]. However, combustion experiments are more complex and involves explosive chemicals, causing more time to adopt. Another interesting field of AI is active learning, which can be used to plan and design expensive combustion experiments [138]. Data visualization is another branch that has largely evolved due to the advent to CNNs, and this is being used for the determination of several phenomenon in combustion [139,140]. To sum up, AI is being crossed-over into many other fields which are already being used in a diverse combustion field. Therefore, as each of the other fields evolve, the innovations in AI will eventually be adapted into combustion.

Acknowledgments

The authors thank Nursulu Kuzhagaliyeva and Francesco Tutino for useful discussions.

References

[1] C. Nieto-Draghi, G. Fayet, B. Creton, X. Rozanska, P. Rotureau, J.-C. de Hemptinne, et al., A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes, Chem. Rev. 115 (24) (2015) 13093–164.

[2] J. George, G. Hautier, Chemist versus Machine: Traditional Knowledge versus Machine Learning Techniques, Trends Chem. 3 (2) (2021) 86–95.

[3] P.P. Plehiers, S.H. Symoens, I. Amghizar, G.B. Marin, C. V Stevens, K.M. Van Geem, Artificial Intelligence in Steam Cracking Modeling: A Deep Learning Algorithm for Detailed Effluent Prediction, Engineering 5 (6) (2019) 1027–40.

[4] S.M.M. Sarathy, A. Farooq, G.T.G.T. Kalghatgi, Recent progress in gasoline surrogate fuels, Prog. Energy Combust. Sci. 65 (2018) 1–42.

[5] C. Ashraf, A. Jain, Y. Xuan, A.C.T. van Duin, ReaxFF based molecular dynamics simulations of ignition front propagation in hydrocarbon/oxygen mixtures under high temperature and pressure conditions, Phys. Chem. Chem. Phys. 19 (7) (2017) 5004–17.

[6] S. Han, X. Li, L. Guo, H. Sun, M. Zheng, W. Ge, Refining Fuel Composition of RP-3 Chemical Surrogate Models by Reactive Molecular Dynamics and Machine Learning, Energy & Fuels 34 (9) (2020) 11381–94.

[7] J. Hachmann, M.A.F. Afzal, M. Haghighatlari, Y. Pal, Building and deploying a cyberinfrastructure for the data-driven design of chemical systems and the exploration of chemical space, Mol. Simul. 44 (11) (2018) 921–9.

[8] A. König, W. Marquardt, A. Mitsos, J. Viell, M. Dahmen, Integrated design of renewable fuels and their production processes: recent advances and challenges, Curr. Opin. Chem.

Eng. 27 (2020) 45–50.

[9] S. Hada, C.C. Solvason, M.R. Eden, Characterization-Based Molecular Design of Bio-Fuel Additives Using Chemometric and Property Clustering Techniques, Front. Energy Res. 2 (2014) 20.

[10] D. Gschwend, A Systematic Search for Next Generation Transportation Fuels PhD Thesis, ETH Zurich, 2018, .

[11] B. Heuser, F. Kremer, S. Pischinger, J. Julis, W. Leitner, Optimization of Diesel Combustion and Emissions with Newly Derived Biogenic Alcohols, SAE/KSAE 2013 International Powertrains, Fuels & Lubricants Meeting, SAE InternationalIn, 2013, .

[12] M. Dahmen, W. Marquardt, Model-Based Formulation of Biofuel Blends by Simultaneous Product and Pathway Design, Energy & Fuels 31 (4) (2017) 4096–121.

[13] P. Gantzer, B. Creton, C. Nieto-Draghi, Inverse-QSPR for de novo Design: A Review, Mol. Inform. 39 (4) (2020) 1900087.

[14] E.O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, A. Aspuru-Guzik, What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery, Annu. Rev. Mater. Res. 45 (1) (2015) 195–216.

[15] B. Burger, P.M. Maffettone, V. V Gusev, C.M. Aitchison, Y. Bai, X. Wang, et al., A mobile robotic chemist, Nature 583 (7815) (2020) 237–41.

[16] D. Delhaye, F.-X. Ouf, D. Ferry, I.K. Ortega, O. Penanhoat, S. Peillon, et al., The MERMOSE project: Characterization of particulate matter emissions of a commercial aircraft engine, J. Aerosol Sci. 105 (2017) 48–63.

[17] P.S. Gromski, A.B. Henson, J.M. Granda, L. Cronin, How to explore chemical space using algorithms and automation, Nat. Rev. Chem. 3 (2) (2019) 119–28.

[18] No Title, Available from: <www.biofuelsflightpath.eu/images/events/2018-nov-12/7_181112_Flightpath_JETSCREEN_overview.pdf>.

[19] C. Jooß, F. Welter, I. Leisten, A. Richert, A.K. Schaar, A.C. Valdez, et al., Scientific Cooperation Engineering in the Cluster of Excellence Integrative Production Technology for High-Wage Countries at RWTH Aachen University, Jeschke S, Isenhardt I, Hees F, Henning K, editors., Automation, Communication and Cybernetics in Science and Engineering 2013/2014, Springer International Publishing, ChamIn, 2014, p. 103–9.

[20] F. Hoppe, B. Heuser, M. Thewes, F. Kremer, S. Pischinger, M. Dahmen, et al., Tailor-made fuels for future engine concepts, Int. J. Engine Res. 17 (1) (2015) 16–27.

[21] J.P. Szybist, S. Busch, R.L. McCormick, J.A. Pihl, D.A. Splitter, M.A. Ratcliff, et al., What fuel properties enable higher thermal efficiency in spark-ignited engines?, Prog. Energy Combust. Sci. 82 (2021) 100876.

[22] G. Fayet, P. Rotureau, How to use QSPR-type approaches to predict properties in the context of Green Chemistry, Biofuels, Bioprod. Biorefining 10 (6) (2016) 738–52.

[23] K.H. Bleicher, H.-J. Böhm, K. Müller, A.I. Alanine, Hit and lead generation: beyond high-throughput screening, Nat. Rev. Drug Discov. 2 (5) (2003) 369–78.

[24] C. Gertig, L. Fleitmann, C. Hemprich, J. Hense, A. Bardow, K. Leonhard, Integrated In Silico Design of Catalysts and Processes based on Quantum Chemistry, Computer Aided Chemical EngineeringIn, 2020, p. 889–94.

[25] D. Flórez-Orrego, J.A.M. Silva, S. de Oliveira Jr., Exergy and environmental comparison of the end use of vehicle fuels: The Brazilian case, Energy Convers. Manag. 100 (2015) 220–31.

[26] S. de Oliveira Junior, Exergy: production, cost and renewability, Springer Science & Business Media, 2012, .

[27] J.-F. Portha, S. Louret, M.-N. Pons, J.-N. Jaubert, Estimation of the environmental impact of a petrochemical process using coupled LCA and exergy analysis, Resour. Conserv. Recycl. 54 (5) (2010) 291–8.

[28] A.P. Mayol, J.L.G.S. Juan, E. Sybingco, A. Bandala, E. Dadios, A.T. Ubando, et al., Environmental impact prediction of microalgae to biofuels chains using artificial intelligence: A life cycle perspective, {IOP} Conf. Ser. Earth Environ. Sci. 463 (2020) 12011.

[29] L. Breiman, Random Forests, Mach. Learn. 45 (1) (2001) 5–32.

[30] D.A. Saldana, L. Starck, P. Mougin, B. Rousseau, B. Creton, On the rational formulation of alternative fuels: melting point and net heat of combustion predictions for fuel compounds using machine learning methods, SAR QSAR Environ. Res. 24 (4) (2013) 259–77.

[31] J.T. Leonard, K. Roy, On Selection of Training and Test Sets for the Development of Predictive QSAR models, QSAR Comb. Sci. 25 (3) (2006) 235–51.

[32] S. Yuan, Z. Zhang, Y. Sun, J.S.-I. Kwon, C. V Mashuga, Liquid flammability ratings predicted by machine learning considering aerosolization, J. Hazard. Mater. 386 (2020) 121640.

[33] I.T. Jolliffe, A Note on the Use of Principal Components in Regression, J. R. Stat. Soc. Ser. C (Applied Stat. 31 (3) (1982) 300–3.

[34] D.A. Saldana, L. Starck, P. Mougin, B. Rousseau, L. Pidol, N. Jeuland, et al., Flash Point and Cetane Number Predictions for Fuel Compounds Using Quantitative Structure Property Relationship (QSPR) Methods, Energy & Fuels 25 (9) (2011) 3900–8.

[35] Y. An, W. Sherman, S.L. Dixon, Kernel-Based Partial Least Squares: Application to Fingerprint-Based QSAR with Model Visualization, J. Chem. Inf. Model. 53 (9) (2013) 2312–21.

[36] Y. Shao, R.S. Lunetta, Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points, ISPRS J. Photogramm. Remote Sens. 70 (2012) 78–87.

[37] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, .

[38] J.C. Hoskins, D.M. Himmelblau, Artificial neural network models of knowledge representation in chemical engineering, Comput. Chem. Eng. 12 (9) (1988) 881–90.

[39] Y. Roh, G. Heo, S.E. Whang, A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective, IEEE Trans. Knowl. Data Eng. (2019) 1.

[40] A.O. Barradas Filho, A.K.D. Barros, S. Labidi, I.M.A. Viegas, D.B. Marques, A.R.S. Romariz, et al., Application of artificial neural networks to predict viscosity, iodine value and induction period of biodiesel focused on the study of oxidative stability, Fuel 145 (2015) 127–35.

[41] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document

recognition, Proc. IEEE 86 (11) (1998) 2278–324.

[42] B. Sattarov, I.I. Baskin, D. Horvath, G. Marcou, E.J. Bjerrum, A. Varnek, De Novo Molecular Design by Combining Deep Autoencoder Recurrent Neural Networks with Generative Topographic Mapping, J. Chem. Inf. Model. 59 (3) (2019) 1182–96.

[43] K. Hanaoka, Deep Neural Networks for Multicomponent Molecular Systems, ACS Omega 5 (33) (2020) 21042–53.

[44] R.S. Bohacek, C. McMartin, W.C. Guida, The art and practice of structure-based drug design: A molecular modeling perspective, Med. Res. Rev. 16 (1) (1996) 3–50.

[45] R. Davis, P. John, Application of Taguchi-based design of experiments for industrial chemical processes, Stat. Approaches with Emphas. Des. Exp. Appl. to Chem. Process. 137 (2018) 137–55.

[46] R.H. Myers, D.C. Montgomery, Response Surface Methodology: Process and Product in Optimization Using Designed Experiments, 1st ed., John Wiley & Sons, Inc., USA, 1995, .

[47] R. van Deursen, J.-L. Reymond, Chemical Space Travel, ChemMedChem 2 (5) (2007) 636–40.

[48] D. Hoksza, D. Svozil, Exploration of Chemical Space by Molecular Morphing, 2011 IEEE 11th International Conference on Bioinformatics and BioengineeringIn, 2011, p. 201–8.

[49] A.M. Virshup, J. Contreras-García, P. Wipf, W. Yang, D.N. Beratan, Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds, J. Am. Chem. Soc. 135 (19) (2013) 7296–303.

[50] C.Y. Cheng, J.E. Campbell, G.M. Day, Evolutionary chemical space exploration for functional materials: computational organic semiconductor discovery, Chem. Sci. 11 (19) (2020) 4922–33.

[51] H.J. Morowitz, J.D. Kostelnik, J. Yang, G.D. Cody, The origin of intermediary metabolism, Proc. Natl. Acad. Sci. 97 (14) (2000) 7704–8.

[52] A.K. Burnham, Van Krevelen Diagrams BT - Encyclopedia of Petroleum Geoscience, Sorkhabi R, editor., Springer International Publishing, ChamIn, 2018, p. 1–5.

[53] C.J. Churchwell, M.D. Rintoul, S. Martin, D.P. Visco, A. Kotu, R.S. Larson, et al., The signature molecular descriptor: 3. Inverse-quantitative structure–activity relationship of ICAM-1 inhibitory peptides, J. Mol. Graph. Model. 22 (4) (2004) 263–73.

[54] M. Popova, O. Isayev, A. Tropsha, Deep reinforcement learning for de novo drug design, Sci. Adv. 4 (7) (2018).

[55] R. Gómez-Bombarelli, J.N. Wei, D. Duvenaud, J.M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, et al., Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules, ACS Cent. Sci. 4 (2) (2018) 268–76.

[56] M. Olivecrona, T. Blaschke, O. Engkvist, H. Chen, Molecular de-novo design through deep reinforcement learning, J. Cheminform. 9 (1) (2017) 48.

[57] E.N. Muratov, E. V Varlamova, A.G. Artemenko, P.G. Polishchuk, V.E. Kuz'min, Existing and Developing Approaches for QSAR Analysis of Mixtures, Mol. Inform. 31 (3-4) (2012) 202–21.

[58] D.A. Saldana, L. Starck, P. Mougin, B. Rousseau, B. Creton, Prediction of Flash Points for Fuel Mixtures Using Machine Learning and a Novel Equation, Energy & Fuels 27 (7) (2013) 3811–20.

[59] G. Cai, Z. Liu, L. Zhang, Q. Shi, S. Zhao, C. Xu, Systematic performance evaluation of gasoline molecules based on quantitative structure-property relationship models, Chem. Eng. Sci. 229 (2021) 116077.

[60] W.L. Kubic, R.W. Jenkins, C.M. Moore, T.A. Semelsberger, A.D. Sutton, Artificial Neural Network Based Group Contribution Method for Estimating Cetane and Octane Numbers of Hydrocarbons and Oxygenated Organic Compounds, Ind. Eng. Chem. Res. 56 (42) (2017) 12236–45.

[61] Z. Liu, L. Zhang, A. Elkamel, D. Liang, S. Zhao, C. Xu, et al., Multiobjective Feature Selection Approach to Quantitative Structure Property Relationship Models for Predicting the Octane Number of Compounds Found in Gasoline, Energy & Fuels 31 (6) (2017) 5828–39.

[62] Z.J. Buras, C. Safta, J. Zádor, L. Sheps, Simulated production of OH, HO2, CH2O, and CO2 during dilute fuel oxidation can predict 1st-stage ignition delays, Combust. Flame 216 (2020) 472–84.

[63] F. vom Lehn, B. Brosius, R. Broda, L. Cai, H. Pitsch, Using machine learning with target-specific feature sets for structure-property relationship modeling of octane numbers and octane sensitivity, Fuel 281 (2020) 118772.

[64] A.M. Schweidtmann, J.G. Rittig, A. König, M. Grohe, A. Mitsos, M. Dahmen, Graph Neural Networks for Prediction of Fuel Ignition Quality, Energy & Fuels 34 (9) (2020) 11395–407.

[65] T. Kessler, E.R. Sacia, A.T. Bell, J.H. Mack, Artificial neural network based predictions of cetane number for furanic biofuel additives, Fuel 206 (2017) 171–9.

[66] L.S. Whitmore, R.W. Davis, R.L. McCormick, J.M. Gladden, B.A. Simmons, A. George, et al., BioCompoundML: A General Biofuel Property Screening Tool for Biological Molecules Using Random Forest Classifiers, Energy & Fuels 30 (10) (2016) 8410–8.

[67] W. Han, Z. Sun, A. Scholtissek, C. Hasse, Machine Learning of ignition delay times under dual-fuel engine conditions, Fuel 288 (2021) 119650.

[68] T.A. Albahri, Structural Group Contribution Method for Predicting the Octane Number of Pure Hydrocarbon Liquids, Ind. Eng. Chem. Res. 42 (3) (2003) 657–62.

[69] A.R. Katritzky, S.H. Slavov, D.A. Dobchev, M. Karelson, Rapid QSPR model development technique for prediction of vapor pressure of organic compounds, Comput. Chem. Eng. 31 (9) (2007) 1123–30.

[70] H.E. McClelland, P.C. Jurs, Quantitative Structure−Property Relationships for the Prediction of Vapor Pressures of Organic Compounds from Molecular Structures, J. Chem. Inf. Comput. Sci. 40 (4) (2000) 967–75.

[71] J. Gao, X. Wang, X. Yu, X. Li, H. Wang, Calculation of polyamides melting point by quantum-chemical method and BP artificial neural networks, J. Mol. Model. 12 (4) (2006) 521–7.

[72] L.H. Hall, C.T. Story, Boiling Point and Critical Temperature of a Heterogeneous Data Set: QSAR with Atom Type Electrotopological State Indices Using Artificial Neural Networks, J. Chem. Inf. Comput. Sci. 36 (5) (1996) 1004–14.

[73] G. Li, Z. Hu, F. Hou, X. Li, L. Wang, X. Zhang, Machine learning enabled high-throughput screening of hydrocarbon molecules for the design of next generation fuels, Fuel 265 (2020) 116968.

[74] Y. Su, Z. Wang, S. Jin, W. Shen, J. Ren, M.R. Eden, An architecture of deep learning in QSPR modeling for the prediction of critical properties using molecular signatures, AIChE J. 65 (9) (2019) e16678.

[75] F. Jirasek, R.A.S. Alves, J. Damay, R.A. Vandermeulen, R. Bamler, M. Bortz, et al., Machine Learning in Thermodynamics: Prediction of Activity Coefficients by Matrix Completion, J. Phys. Chem. Lett. 11 (3) (2020) 981–5.

[76] D.A. Saldana, L. Starck, P. Mougin, B. Rousseau, N. Ferrando, B. Creton, Prediction of Density and Viscosity of Biofuel Compounds Using Machine Learning Methods, Energy & Fuels 26 (4) (2012) 2416–26.

[77] G. Cai, Z. Liu, L. Zhang, S. Zhao, C. Xu, Quantitative Structure–Property Relationship Model for Hydrocarbon Liquid Viscosity Prediction, Energy & Fuels 32 (3) (2018) 3290–8.

[78] S. Kosir, J. Heyne, J. Graham, A machine learning framework for drop-in volume swell characteristics of sustainable aviation fuel, Fuel 274 (2020) 117832.

[79] A. Sanaeifar, A. Jafari, Determination of the oxidative stability of olive oil using an integrated system based on dielectric spectroscopy and computer vision, Inf. Process. Agric. 6 (1) (2019) 20–5.

[80] J. Frutiger, C. Marcarie, J. Abildskov, G. Sin, A Comprehensive Methodology for Development, Parameter Estimation, and Uncertainty Analysis of Group Contribution Based Property Models—An Application to the Heat of Combustion, J. Chem. Eng. Data 61 (1) (2016) 602–13.

[81] K.K. Yalamanchi, V.C.O. van Oudenhoven, F. Tutino, M. Monge-Palacios, A. Alshehri, X. Gao, et al., Machine Learning To Predict Standard Enthalpy of Formation of Hydrocarbons, J. Phys. Chem. A 123 (38) (2019) 8305–13.

[82] Y. Pan, J.C. Jiang, R. Wang, J.J. Jiang, Predicting the net heat of combustion of organic compounds from molecular structures based on ant colony optimization, J. Loss Prev. Process Ind. 24 (1) (2011) 85–9.

[83] A. Sosnowska, M. Barycki, K. Jagiello, M. Haranczyk, A. Gajewicz, T. Kawai, et al., Predicting enthalpy of vaporization for Persistent Organic Pollutants with Quantitative Structure–Property Relationship (QSPR) incorporating the influence of temperature on volatility, Atmos. Environ. 87 (2014) 10–8.

[84] K.K. Yalamanchi, M. Monge-Palacios, V.C.O. van Oudenhoven, X. Gao, S.M. Sarathy, Data Science Approach to Estimate Enthalpy of Formation of Cyclic Hydrocarbons, J. Phys. Chem. A 124 (31) (2020) 6270–6.

[85] M.N. Aldosari, K.K. Yalamanchi, X. Gao, S.M. Sarathy, Predicting entropy and heat capacity of hydrocarbons using machine learning, Energy AI 4 (2021) 100054.

[86] D.D. Das, P.C. St. John, C.S. McEnally, S. Kim, L.D. Pfefferle, Measuring and predicting sooting tendencies of oxygenates, alkanes, alkenes, cycloalkanes, and aromatics on a unified scale, Combust. Flame 190 (2018) 349–64.

[87] A. Smith, A. Keane, J.A. Dumesic, G.W. Huber, V.M. Zavala, A machine learning framework for the analysis and prediction of catalytic activity from experimental data,

Appl. Catal. B Environ. 263 (2020) 118257.

[88] I. Takigawa, K. Shimizu, K. Tsuda, S. Takakusagi, Machine Learning Predictions of Factors Affecting the Activity of Heterogeneous Metal Catalysts, Tanaka I, editor., Nanoinformatics, Springer Singapore, SingaporeIn, 2018, p. 45–64.

[89] Y. Pan, J. Jiang, R. Wang, H. Cao, J. Zhao, Quantitative Structure–Property Relationship Studies for Predicting Flash Points of Organic Compounds using Support Vector Machines, QSAR Comb. Sci. 27 (8) (2008) 1013–9.

[90] J.A. Lazzús, Prediction of flammability limit temperatures from molecular structures using a neural network–particle swarm algorithm, J. Taiwan Inst. Chem. Eng. 42 (3) (2011) 447–53.

[91] S. Yuan, Z. Jiao, N. Quddus, J.S.-I. Kwon, C. V Mashuga, Developing Quantitative Structure–Property Relationship Models To Predict the Upper Flammability Limit Using Machine Learning, Ind. Eng. Chem. Res. 58 (8) (2019) 3531–7.

[92] L. Carlsson, E.A. Helgee, S. Boyer, Interpretation of Nonlinear QSAR Models Applied to Ames Mutagenicity Data, J. Chem. Inf. Model. 49 (11) (2009) 2551–8.

[93] C.G. BERTINETTO, PREDICTION OF THE PHYSICO-CHEMICAL PROPERTIES OF LOW AND HIGH MOLECULAR WEIGHT COMPOUNDS (2010).

[94] Z. Jiao, P. Hu, H. Xu, Q. Wang, Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications, ACS Chem. Heal. Saf. 27 (6) (2020) 316–34.

[95] J.H. Al-Fahemi, N.A. Albis, E.A.M. Gad, QSPR Models for Octane Number Prediction, Sajan D, editor., J. Theor. Chem. 2014 (2014) 520652.

[96] G. Mendes, H.G. Aleme, P.J.S. Barbeira, Determination of octane numbers in gasoline by distillation curves and partial least squares regression, Fuel 97 (2012) 131–6.

[97] J.M. Andrade, S. Muniategui, D. Prada, Prediction of clean octane numbers of catalytic reformed naphthas using FT-m.i.r. and PLS, Fuel 76 (11) (1997) 1035–42.

[98] S.R. Daly, K.E. Niemeyer, W.J. Cannella, C.L. Hagen, Predicting fuel research octane number using Fourier-transform infrared absorption spectra of neat hydrocarbons, Fuel 183 (2016) 359–65.

[99] R. Palani, A. AbdulGani, N. Balasubramanian, Treatment of Tannery Effluent Using a Rotating Disc Electrochemical Reactor, Water Environ. Res. 89 (1) (2017) 77–85.

[100] A.G. Abdul Jameel, Y. Han, O. Brignoli, S. Telalović, A.M. Elbaz, H.G. Im, et al., Heavy fuel oil pyrolysis and combustion: Kinetics and evolved gases investigated by TGA-FTIR, J. Anal. Appl. Pyrolysis 127 (2017) 183–95.

[101] J.M. de Paulo, J.E.M. Barros, P.J.S. Barbeira, A PLS regression model using flame spectroscopy emission for determination of octane numbers in gasoline, Fuel 176 (2016) 216–21.

[102] A.G. Abdul Jameel, A.M. Elbaz, A.-H. Emwas, W.L. Roberts, S.M. Sarathy, Calculation of Average Molecular Parameters, Functional Groups, and a Surrogate Molecule for Heavy Fuel Oils Using 1H and 13C Nuclear Magnetic Resonance Spectroscopy, Energy & Fuels 30 (5) (2016) 3894–905.

[103] A.G. Abdul Jameel, S.M. Sarathy, Prediction of RON and MON of gasoline-ethanol using 1 H NMR spectroscopy, Proc. Eur. Combust. Meet (2017).

[104] A.G. Abdul Jameel, V. Van Oudenhoven, A.-H. Emwas, S.M. Sarathy, Predicting Octane Number Using Nuclear Magnetic Resonance Spectroscopy and Artificial Neural Networks, Energy & Fuels 32 (5) (2018) 6309–29.

[105] P.E. Flecher, W.T. Welch, S. Albin, J.B. Cooper, Determination of octane numbers and Reid vapor pressure in commercial gasoline using dispersive fiber-optic Raman spectroscopy, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 53 (2) (1997) 199–206.

[106] L. Guan, X.L. Feng, Z.C. Li, G.M. Lin, Determination of octane numbers for clean gasoline using dielectric spectroscopy, Fuel 88 (8) (2009) 1453–9.

[107] X. Shi, H. Li, Z. Song, X. Zhang, G. Liu, Quantitative composition-property relationship of aviation hydrocarbon fuel based on comprehensive two-dimensional gas chromatography with mass spectrometry and flame ionization detector, Fuel 200 (2017) 395–406.

[108] R. Van de Vijver, J. Zádor, KinBot: Automated stationary point search on potential energy surfaces, Comput. Phys. Commun. 248 (2020) 106947.

[109] J.S. Schreck, C.W. Coley, K.J.M. Bishop, Learning Retrosynthetic Planning through Simulated Experience, ACS Cent. Sci. 5 (6) (2019) 970–81.

[110] P.L. Bhoorasingh, R.H. West, Transition state geometry prediction using molecular group contributions, Phys. Chem. Chem. Phys. 17 (48) (2015) 32173–82.

[111] L. Pattanaik, J.B. Ingraham, C.A. Grambow, W.H. Green, Generating transition states of isomerization reactions with deep learning, Phys. Chem. Chem. Phys. 22 (41) (2020) 23618–26.

[112] K.T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, R.J. Maurer, Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions, Nat. Commun. 10 (1) (2019) 5024.

[113] J.S. Smith, B.T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, et al., Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning, Nat. Commun. 10 (1) (2019) 2903.

[114] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O.A. von Lilienfeld, K.-R. Müller, et al., Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space, J. Phys. Chem. Lett. 6 (12) (2015) 2326–31.

[115] L.-P. Wang, A. Titov, R. McGibbon, F. Liu, V.S. Pande, T.J. Martínez, Discovering chemistry with an ab initio nanoreactor, Nat. Chem. 6 (12) (2014) 1044–8.

[116] M. Döntgen, M.-D. Przybylski-Freund, L.C. Kröger, W.A. Kopp, A.E. Ismail, K. Leonhard, Automated Discovery of Reaction Pathways, Rate Constants, and Transition States Using Reactive Molecular Dynamics Simulations, J. Chem. Theory Comput. 11 (6) (2015) 2517–24.

[117] P. Yoo, M. Sakano, S. Desai, M.M. Islam, P. Liao, A. Strachan, Neural network reactive force field for C, H, N, and O systems, npj Comput. Mater. 7 (1) (2021) 9.

[118] A.P. Bartók, S. De, C. Poelking, N. Bernstein, J.R. Kermode, G. Csányi, et al., Machine learning unifies the modeling of materials and molecules, Sci. Adv. 3 (12) (2017) e1701816.

[119] T. Toyao, Z. Maeno, S. Takakusagi, T. Kamachi, I. Takigawa, K. Shimizu, Machine

Learning for Catalysis Informatics: Recent Applications and Prospects, ACS Catal. 10 (3) (2020) 2260–97.

[120] H.E. Reynel-Ávila, A. Bonilla-Petriciolet, J.C. Tapia-Picazo, An artificial neural network-based NRTL model for simulating liquid-liquid equilibria of systems present in biofuels production, Fluid Phase Equilib. 483 (2019) 153–64.

[121] C.J. Lapeyre, A. Misdariis, N. Cazard, D. Veynante, T. Poinsot, Training convolutional neural networks to estimate turbulent sub-grid scale reaction rates, Combust. Flame 203 (2019) 255–64.

[122] L. Pulga, G.M. Bianchi, S. Falfari, C. Forte, A machine learning methodology for improving the accuracy of laminar flame simulations with reduced chemical kinetics mechanisms, Combust. Flame 216 (2020) 72–81.

[123] R. Ranade, S. Alqahtani, A. Farooq, T. Echekki, An extended hybrid chemistry framework for complex hydrocarbon fuels, Fuel 251 (2019) 276–84.

[124] A.J. Sharma, R.F. Johnson, D.A. Kessler, A. Moses, Deep Learning for Scalable Chemical Kinetics, AIAA Scitech 2020 Forum, American Institute of Aeronautics and AstronauticsIn, 2020, (AIAA SciTech Forum). .

[125] J. An, H. Wang, B. Liu, K.H. Luo, F. Qin, G.Q. He, A deep learning framework for hydrogen-fueled turbulent combustion simulation, Int. J. Hydrogen Energy 45 (35) (2020) 17992–8000.

[126] M. Bracconi, M. Maestri, Training set design for machine learning techniques applied to the approximation of computationally intensive first-principles kinetic models, Chem. Eng. J. 400 (2020) 125469.

[127] C. März, J. Werfel, J. Kühne, R. Scholz, Approaches for a New Generation of Fast-Computing Catalyst Models, Emiss. Control Sci. Technol. 6 (2) (2020) 254–68.

[128] L.Y. Pfund, A.J. Matzger, Towards Exhaustive and Automated High-Throughput Screening for Crystalline Polymorphs, ACS Comb. Sci. 16 (7) (2014) 309–13.

[129] A.P. Toropova, A.A. Toropov, QSPR and nano-QSPR: What is the difference?, J. Mol. Struct. 1182 (2019) 141–9.

[130] P. Needham, Is water a mixure? Bridging the distinction between physical and chemical properties, Stud. Hist. Philos. Sci. Part A 39 (1) (2008) 66–77.

[131] A.M. TURING, I.—COMPUTING MACHINERY AND INTELLIGENCE, Mind LIX (236) (1950) 433–60.

[132] M. Shanahan, M. Crosby, B. Beyret, L. Cheke, Artificial Intelligence and the Common Sense of Animals., Trends Cogn. Sci. 24 (11) (2020) 862–72.

[133] C.W. Coley, N.S. Eyke, K.F. Jensen, Autonomous Discovery in the Chemical Sciences Part II: Outlook, Angew. Chemie Int. Ed. 59 (52) (2020) 23414–36.

[134] D. Lowe, AI designs organic syntheses., Nature 555 (7698) (2018) 592–3.

[135] L. Cao, D. Russo, K. Felton, D. Salley, A. Sharma, G. Keenan, et al., Optimization of Formulations Using Robotic Experiments Driven by Machine Learning DoE, Cell Reports Phys. Sci. 2 (1) (2021) 100295.

[136] J.M. Granda, L. Donina, V. Dragone, D.-L. Long, L. Cronin, Controlling an organic synthesis robot with machine learning to search for new reactivity, Nature 559 (7714)

(2018) 377–81.

[137] A.D. Clayton, J.A. Manson, C.J. Taylor, T.W. Chamberlain, B.A. Taylor, G. Clemens, et al., Algorithms for the self-optimisation of chemical reactions, React. Chem. Eng. 4 (9) (2019) 1545–54.

[138] X.E. Pantazi, D. Moshou, D. Kateris, I. Gravalos, P. Xyradakis, Automatic Identification of Gasoline – Biofuel Blend Type in an Internal Combustion Four-stroke Engine based on Unsupervised Novelty Detection and Active Learning, Procedia Technol. 8 (2013) 229–37.

[139] Z. Han, J. Li, B. Zhang, M.M. Hossain, C. Xu, Prediction of combustion state through a semi-supervised learning model and flame imaging, Fuel 289 (2021) 119745.

[140] A. Hanuschkin, S. Zündorf, M. Schmidt, C. Welch, J. Schorr, S. Peters, et al., Investigation of cycle-to-cycle variations in a spark-ignition engine based on a machine learning analysis of the early flame kernel, Proc. Combust. Inst. (2020).