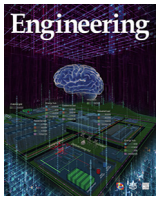
Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
AI Energizes Process Manufacturing—Perspective

Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats

Maarten R. Dobbelaere [a], Pieter P. Plehiers [a], Ruben Van de Vijver [a], Christian V. Stevens [b], Kevin M. Van Geem [b,*]

[a] Laboratory for Chemical Technology, Department of Materials, Textiles and Chemical Engineering, Ghent University, Ghent 9052, Belgium
[b] SynBioC Research Group, Department of Green Chemistry and Technology, Faculty of Bioscience Engineering, Ghent University, Ghent 9000, Belgium

ARTICLE INFO

Article history:
Received 16 October 2020
Revised 16 January 2021
Accepted 22 March 2021
Available online 29 July 2021

Keywords:
Artificial intelligence
Machine learning
Reaction engineering
Process engineering

ABSTRACT

Chemical engineers rely on models for design, research, and daily decision-making, often with potentially large financial and safety implications. Previous efforts a few decades ago to combine artificial intelligence and chemical engineering for modeling were unable to fulfill the expectations. In the last five years, the increasing availability of data and computational resources has led to a resurgence in machine learning-based research. Many recent efforts have facilitated the roll-out of machine learning techniques in the research field by developing large databases, benchmarks, and representations for chemical applications and new machine learning frameworks. Machine learning has significant advantages over traditional modeling techniques, including flexibility, accuracy, and execution speed. These strengths also come with weaknesses, such as the lack of interpretability of these black-box models. The greatest opportunities involve using machine learning in time-limited applications such as real-time optimization and planning that require high accuracy and that can build on models with a self-learning ability to recognize patterns, learn from data, and become more intelligent over time. The greatest threat in artificial intelligence research today is inappropriate use because most chemical engineers have had limited training in computer science and data analysis. Nevertheless, machine learning will definitely become a trustworthy element in the modeling toolbox of chemical engineers.

1. Introduction

In 130 years of chemical engineering, mathematical modeling has been invaluable to engineers for understanding and designing chemical processes. Octave Levenspiel even stated that modeling stands out as the primary development in chemical engineering [1]. Today, in a fast-moving world, there are more challenges than ever. The ability to predict the outcomes of certain events is necessary, regardless of whether such events are related to the discovery and synthesis of active pharmaceutical ingredients for new diseases or to improvements in process efficiencies to meet stricter environmental legislation. These events range from the reaction rate of a surface reaction or the selectivity of a reaction in a reactor, to the control of the heat supply to that reactor. Predictions can be made using theoretical models, which have been constructed for centuries. The Navier–Stokes equations [2,3], which describe viscous fluid behavior, are one example of such a theoretical model. However, many of these models cannot be solved analytically for realistic systems and require a considerable amount of computational power to solve numerically. This drawback has ensured that most engineers first use simple models to describe reality. An important historical—yet still relevant—example is Prandtl's boundary layer model [4]. In computational chemistry, scientists and engineers are willing to give up some accuracy in favor of time. This willingness explains the popularity of density functional theory, in comparison with higher-level-of-theory models. However, in many situations, higher accuracy is desired.

Decades of modeling, simulations, and experiments have provided the chemical engineering community with a massive amount of data, which adds the option of making predictions from experience as an extra modeling toolkit. Machine learning models are statistical and mathematical models that can "learn" from experience and discover patterns in data without the need for explicit, rule-based programming. As a field of study, machine

* Corresponding author.
E-mail address: Kevin.VanGeem@UGent.be (K.M. Van Geem).

https://doi.org/10.1016/j.eng.2021.03.019

Please cite this article as: M.R. Dobbelaere, P.P. Plehiers, R. Van de Vijver et al., Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats, Engineering, https://doi.org/10.1016/j.eng.2021.03.019

learning is a subset of artificial intelligence (AI). AI is the ability of machines to perform tasks that are generally linked to the behavior of intelligent beings, such as humans. As shown in Fig. 1, this field is not particularly new. The term "artificial intelligence" was coined at Dartmouth College, USA in 1956, at a summer workshop for mathematicians who aimed at developing more cognizant machines. From that point on, it took more than a decade before the first attempts were made to apply AI in chemical engineering [5]. In the 1980s, greater efforts were made in the field with the use of rule-based expert systems, which are considered to be the simplest forms of AI. By that time, the field of machine learning had started to grow, but in the chemical engineering community, with some exceptions, a lag of about 10 years was experienced in the growth of machine learning. A sudden rise in publications on AI applications in chemical engineering in the 1990s can be observed, with the adoption of clustering algorithms, genetic algorithms, and—most successfully—artificial neural networks (ANNs). Nevertheless, the trend did not persist. Venkatasubramanian [6] names the lack of powerful computing and the difficult task of creating the algorithms as possible causes for this loss in interest.

The past decade marked a breakthrough in deep learning, a subset of machine learning that constructs ANNs to mimic the human brain. As mentioned above, ANNs gained popularity among chemical engineers in the 1990s; however, the difference of the deep learning era is that deep learning provides the computational means to train neural networks with multiple layers—the so-called deep neural networks. These new developments triggered chemical engineers, as reflected by an exponential rise in publications on the topic. In the past, AI techniques could never become a standard tool in chemical engineering; thus, it can be asked whether this is finally the moment. In this perspective article, we will first give an overview of the three major links in machine learning today, applied to chemical engineering. In what follows, the growing potential of machine learning in chemical engineering will be critically discussed; we will examine the pros and cons and list possible reasons for why machine learning in chemical engineering with remain "hot" or end up as a "not."

2. Machine learning ABCs

2.1. The "A" in machine learning ABCs: Data

A machine learning approach consists of three important links, as illustrated in Fig. 2: data, representations, and models. The first link in a machine learning approach is the data that is used to train the model. As will be discussed later, the data used also proves to be the weakest link in the machine learning process. Virtually any dataset containing results from experiments, first-principles calculations, or complex simulation models can be used to train a model. However, because it is expensive to gather large amounts of accurate data, it is customary to make use of "big data" approaches—using large databases from various existing sources. Due to the cost of real experiments, these large quantities of data are usually obtained via fast simulations or text mining from patents and published work. The increased digitalization of research provides the scientific community with a plethora of open-source and commercial databases. Examples of commonly used sources of chemical information are Reaxys [7], SciFinder [8], and ChemSpace [9] for reaction chemistry and properties; GDB-17 [10] for small drug-like molecules; and National Institute of Standards and Technology (NIST) [11] and International Union of Pure and Applied Chemistry (IUPAC) [12] for molecular properties such as solubility. In addition, several benchmarking datasets have been created to enable comparison between different machine learning models. Examples of these benchmarks are QM9 and Alchemy, for quantum chemical properties [13]; and ESOL [14] and FreeSolv [15], for solubilities. Before using any dataset for machine learning-based modeling, several steps should be undertaken to ensure that the used data is of high enough quality. The general aspect of ensuring data quality—from its generation to its storage—is known as data curation. More details about the necessity and consequences of data curation are discussed further on.

Several differences concerning data usage exist between machine learning—and, more specifically, deep learning methods—and traditional modeling. First, ANNs learn from data and

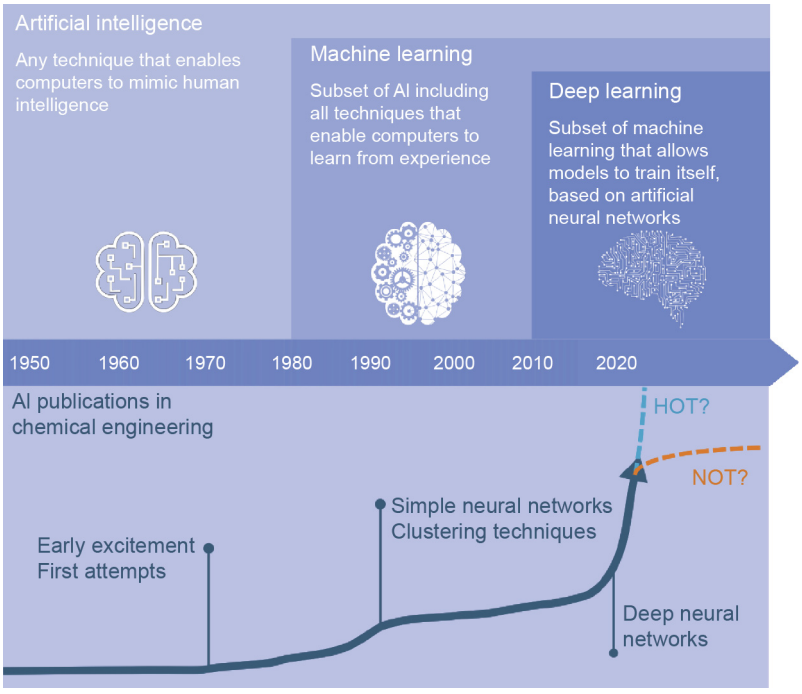


Fig. 1. Timeline of artificial intelligence, machine learning, and deep learning. The evolution of publications about AI in chemical engineering shows that a rise in publications is followed by a phase of disinterest. Currently, AI in chemical engineering is once again in a "hot" phase, and it is unclear whether or not the curve will soon flatten out.

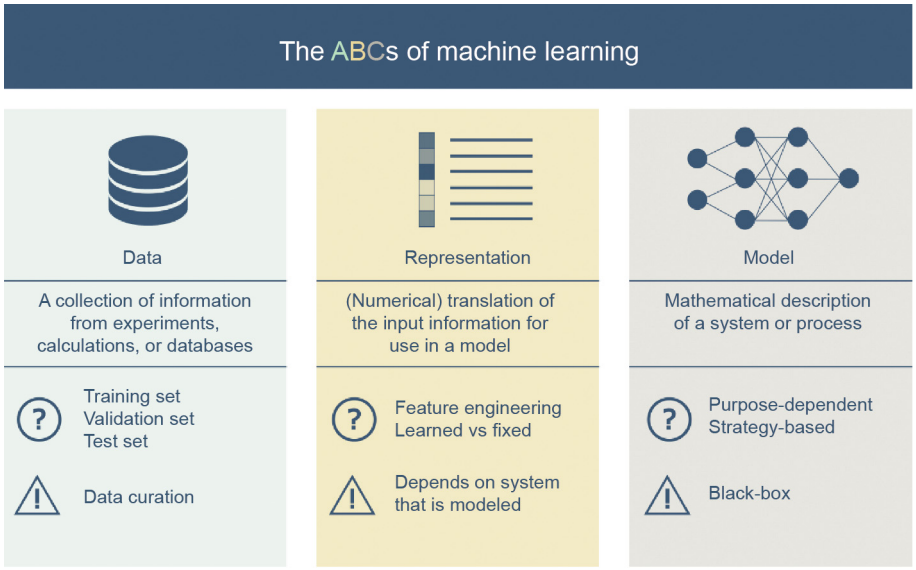


Fig. 2. The three major links in machine learning for chemical engineering; every part has an impact on the eventual prediction performance and should be handled carefully.

train themselves, although doing so requires large amounts of data. Therefore, training datasets generally contain tens to hundreds of thousands of data points. Second, the dataset is split into three instead of two sets: a training, validation, and test set. Both the training and validation sets are used in the training phase, while only the data in the training set is used for fitting. The validation set is an independent dataset that provides an unbiased evaluation of the model fit during the training phase. The test set evaluates the final model fit with unseen data and is generally the main indicator of the model quality.

2.2. The "B" in machine learning ABCs: Representation

A second important link in a machine learning method is how the data is represented in the model. Even when the data is already in numerical format, the selection of the variables or features that will make up the model input can have a significant impact on the model performance. This process is known as feature selection and has been the topic of several studies [16–19]. Limiting the number of selected features may reduce the computational cost of both training and executing the model, while improving the overall accuracy. This feature-selection process is of lesser importance in so-called deep learning methods, which are assumed to internally select those features that are considered to be important [20]. Then, an input layer that consists of basic process parameters (e.g., pressure, temperature, residence time), feed characterizations (e.g., distillation curves, feed compositions), or catalyst properties (e.g., surface area, calcination time) is often sufficient [21–27]. However, the task of representing the data becomes far more challenging in the case of non-numerical data, such as molecules and reactions.

Chemical engineering tasks often involve molecules and/or chemical reactions. Creating suitable numerical representations of these data types is a developing field in itself. In computer applications, the molecular constitution is typically represented by a line-based identifier, such as the simplified molecular-input line-entry system (SMILES) [28] or the (IUPAC) international chemical identifiers (InChIs) [29], or as three-dimensional (3D) coordinates. Recently, self-referencing embedded strings (SELFIES) [30] have been developed as a molecular string representation designed for machine learning applications. The molecular information is translated into a feature vector or tensor that is used as input for a deep neural network or another machine learning model. The first way to represent a molecule is by using a (set of) well-chosen molecular descriptor(s), such as the molecular weight, dipole moment, or dielectric constant [31–33]. Another way to generate a molecular feature vector is by starting from the 3D geometry. Coulomb matrices [34], bags of bonds [35], and histograms of distances, angles, and dihedrals [36] are a few examples of geometry-based representations. However, 3D coordinates or calculated properties are generally unavailable in many applications. In such cases, the representation can be created starting from a molecular graph, resulting in so-called topology-based representations.

In topology-based representations, only a line-based identifier is available. Encoders exist that directly translate the line-based identifier into a representation with techniques from natural language processing [37–41], but usually the line-based identifier is transformed into a feature vector in a similar fashion to geometry-based representations [42–60]. This is done by adding simple atom and bond features to the molecular graph and then transmitting the information iteratively between atoms and bonds. Circular fingerprints [42–46] based on the Morgan algorithm [61], such as the extended-connectivity fingerprint [62], were among the first molecular representations for machine learning applications. These fingerprints are so-called fixed molecular representations because they do not change during the training of the machine learning model. They remain popular in drug design for rapidly predicting the physical, chemical, and biological properties of candidate drugs [63]. Because a fixed representation vector represents a molecule by the same vector in every prediction task, this type of input layer seems to conflict with the definition of a deep neural network, which is assumed to learn the important features [64]. There is a growing tendency to focus on learning how to represent a molecule [47,52] instead of on human-engineering the feature vector, as it is assumed that better capturing of the features will lead to higher accuracy, with less data and at a lower computational cost [53,58].

Learned molecular representations are created as part of the prediction model. Starting from several initial molecular features—such as the heavy atoms, bond types, and ring features—a molecular representation is created that is updated during training. This choice also indicates that a molecule has different representations depending on the prediction task. An extensive variety of learned topology-based representations [47–58] can be

described using the message-passing neural network framework reviewed by Gilmer et al. [59]. The weighted transfer of atom and bond information throughout the molecular graph is characteristic of message-passing neural networks. Many different representations exist, ranging in complexity, but it is important to note that a single representation that works for all kind of molecular properties has not (yet) been developed [65]. For a more detailed overview of the state of the art in representing molecules, readers are referred to the review by David et al. [60].

Chemical reactions are more complex data types than molecules. Similar to line-based molecular identifiers, reactions can be identified by reaction SMILES [66] and reaction InChI (RInChI) [67], whereas SMIRKS [66] identify reaction mechanisms. As for molecules, chemical reactions should also be vectorized in order to be useful in machine learning models. The most straightforward method is to start from the molecular descriptors (e.g., fingerprints) of the reagents and sum [68], subtract [50,69], or concatenate [70–72] them. Another approach is to learn a reaction representation based on the atoms and bonds that take actively part in the reaction [73]. Reactions can also be kept as text (typically InChI) and, with a neural machine translation, the organic reaction product is then considered to be a translation of the reaction products [58,74–78].

2.3. The "C" in machine learning ABCs: Model

The final prerequisite for a machine learning method is a modeling strategy. There is a wide variety of machine learning models to choose from. Models can be categorized in different ways, either by purpose (classification or regression) or by learning methodology (unsupervised, supervised, active, or transfer learning). Generally speaking, the term "machine learning" can be applied to any method in which correlations within datasets are implicitly modeled [79,80]. Therefore, many techniques that are currently referred to as machine learning methods were in use long before they were termed machine learning. Two such examples are Gaussian mixture modeling and principal component analysis (PCA), which originated in, respectively, the late 1800s [81] and the early 1900s [82,83]. Both examples are now regarded as unsupervised machine learning algorithms. Other similar unsupervised clustering methods are t-distributed stochastic neighbor embedding (*t*-SNE) [84] and density-based spatial clustering of applications with noise (DBSCAN) [85]. Fig. 3 shows the difference between unsupervised and supervised learning techniques, with a non-exhaustive list of useful algorithms for a specific task. In unsupervised learning, the algorithm does not need any "solutions" or labels to learn; it will discover patterns by itself. Unsupervised learning techniques have been used for various purposes in chemical engineering. PalkovitsR and PalkovitsS [86] used the *k*-means algorithm [87] for clustering catalysts based on their features and *t*-SNE for the visualization of high-dimensional catalyst representations. Not only used for catalysis, *t*-SNE is the preferred method for visualizing high-dimensional data; it has also been used in the context of fault diagnosis in chemical processes [88,89] and for predicting reaction conditions [69,90]. PCA is another algorithm for reducing dimensionality and has been used multiple times by chemical engineers for determining the features that account for the most variance in the training set [91–97]. In addition, PCA is used for outlier detection [93,98]. Other algorithms used to detect anomalies include DBSCAN and long short-term memory (LSTM) [99,100]. Interested readers are referred to Géron's book [101] for a further introduction to machine learning algorithms.

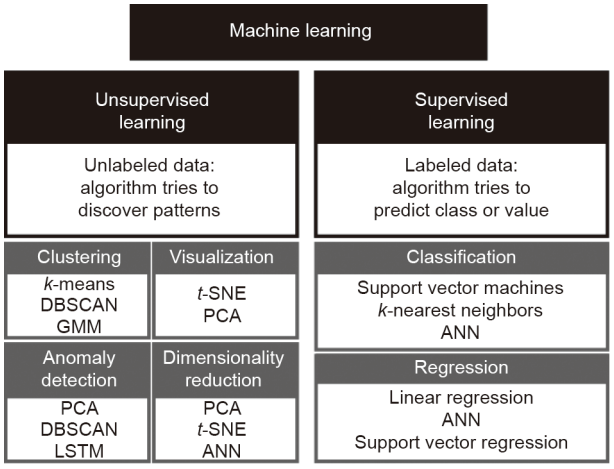


Fig. 3. Overview of unsupervised and supervised machine learning algorithms; a non-exhaustive list of useful algorithms is included. GMM: gaussian mixture modeling; LSTM: long short-term memory; *t*-SNE: *t*-distributed stochastic neighbor embedding.

When the dataset is labeled—that is, when the correct classification of each data point is known—supervised classification methods such as decision trees (and, by extension, random forests) can be used [102,103]. Support vector machines are another possible supervised classification method [104]. Although support vector machines are commonly used for classification purposes, extensions have been made to allow regression via support vector machines as well. Regression problems require supervised or active learning methods, although, in principle, any supervised learning method can be incorporated into an active learning approach. ANNs, and all their possible variations [105–113], are the method that is most commonly associated with machine learning. Depending on the application, one might choose feed-forward ANNs (for feature-based classification or regression), convolutional neural networks (for image processing), or recurrent neural networks (for anomaly detection). A chemical engineer might encounter convolutional neural networks used for representing molecules (see Section 2.2) [42–60] and ANNs [32,33,47,91,114–117], support vector machines [32], or kernel ridge regression [36,118] for predicting the properties of the representations. ANNs have been applied as a black-box modeling tool for numerous applications in catalysis [23], chemical process control [119], and chemical process optimization [120]. A popular algorithm for classifying data points when the labels are known is *k*-nearest neighbors, which has been used, for example, for chemical process monitoring [121,122] and clustering of catalysts [86,123,124].

3. Strengths

In this and the following sections, we give a detailed overview of the strengths, weaknesses, opportunities, and threats in the use of machine learning for chemical engineers. Fig. 4 summarizes what is described in the next sections.

Machine learning techniques have gained popularity in chemistry and chemical engineering for revealing patterns in data that human scientists are unable to discover. In contrast to physical models, which rely explicitly on physical equations (resulting from discovered patterns), machine learning models are not specifically programmed to solve a certain problem. For classification problems, this implies that not a single explicitly defined decision function must be programmed. For regression problems, this implies that no detailed model equations must be derived or parametrized [80]. These advantages allow efficient upscaling to large systems and datasets without the need for extensive computational resources. An example is the current boom in predicting quantum

Fig. 4. Strengths, weaknesses, opportunities, and threats in using machine learning as a modeling tool in chemical engineering.

chemical properties using machine learning [32,33,35–37,39,40,47,49,50,52,55,65,68,71,73,115]. The usual *ab initio* methods often require hours or days to calculate the properties of a single molecule. Well-trained machine learning models can make accurate predictions in a fraction of a second. Of course, other fast techniques that can predict accurately have already been developed, but they are limited in application range compared with machine learning models [125]. The inability to extrapolate is the major weakness of machine learning, but the application range can be extended quite easily by simply adding new data points. Active learning [126,127] makes it possible to expand the range with a minimal amount of new data, which is ideal for cases in which labeling is expensive (i.e., finding the true values of data points), such as quantum chemical calculations [116] or chemical experiments [72,128,129]. Furthermore, existing machine learning models, such as ChemProp [47] and SchNet [130,131], are ready to use and do not require experience. Machine learning in general has become very accessible with packages such as scikit-learn [132] and TensorFlow [133], and frameworks like Keras [134] (now part of TensorFlow [133]) or PyTorch [135], which restrict the training of a deep learning model to just a few lines of code. Such packages and frameworks give scientists the opportunity to shift their focus to the physical meaning of their research instead of spending precious time on developing high-level computer models.

4. Weaknesses

One of the main weaknesses of machine learning approaches is their black-box nature. Given a certain input, the approaches provide an output. This situation is illustrated by Fig. 5. Based on the statistical performance of the model on a test dataset, certain statements can be made about the accuracy and reliability of the generated output. Detailed analysis of the model hyperparameters (e.g., the number of nodes in an ANN) can be tedious, but can provide some insight into the correlations that have been learned by the model. However, extracting physically meaningful explanations for certain behaviors is infeasible. Hence, regardless of their speed and accuracy, machine learning models are a poor modeling choice for explanatory studies.

This lack of interpretability contributes to the difficulty of designing a proper machine learning model. As in any model, a machine learning model can overfit or underfit the data, with the proper model being situated somewhere in between. The risk of overfitting is typically much greater than the risk of underfitting for machine learning models, and depends on the quality and quantity of the training data, and on the complexity of the model. Overfitting is an intrinsic property of the model structure and does not depend on the actual values of the hyperparameters—it can be compared to fitting a (noisy) linear dataset with a polynomial of very high order. In deep learning, overfitting usually manifests itself in the form of overtraining, which arises when the model is shown the same data too many times. This results in the model memorizing noise instead of capturing general patterns. Overtraining can be identified by comparing the model performance on the training data with its performance on the validation and test datasets. If the training performance is much better than the validation performance, the model may be overtrained. Finding the number of training epochs is often a difficult exercise. In order to avoid overfitting, a machine learning model requires a stopping criterion, such as in other optimization problems. In traditional modeling, where models typically involve at least some form of simplification with respect to reality, this stopping criterion is typically based on the change in performance on the training dataset, as achieving a high accuracy of the training data is the main challenge due to the simplifications. Achieving accuracy on the training dataset is typically not the issue for machine learning models; rather, the challenge mainly lies in achieving high accuracy on data the model was not directly trained on. Therefore, the stopping criterion should be based on the performance of the model on "unseen" data—the so-called validation dataset. For rigorously testing the

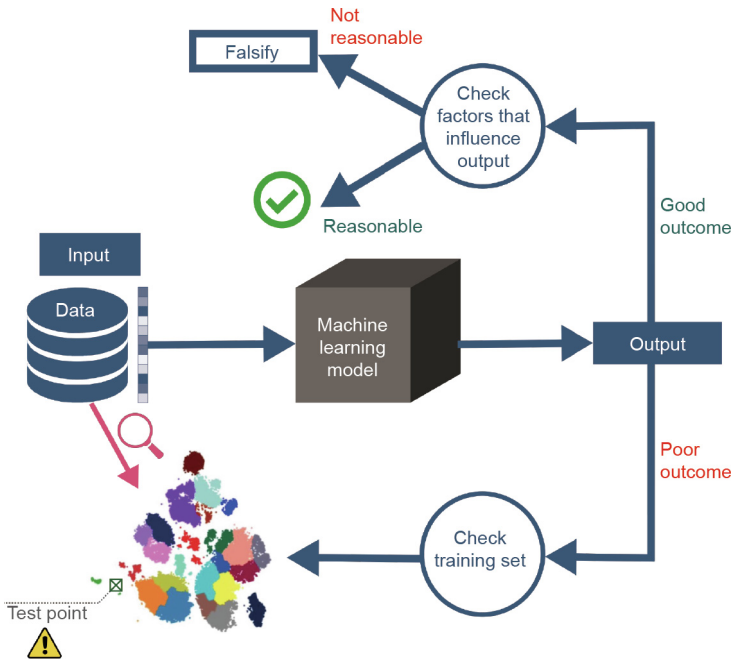


Fig. 5. Unraveling the results from black-box models. A poor result is typically related to the training set used. When testing outside of the application range, a warning signal should be raised. Good results require validation to understand what the model learns.

optimized dataset, a completely independent dataset—the test dataset—is required, as is also common practice in traditional modeling approaches.

A final—but often critical—weakness in machine learning approaches is the data itself that was used. If there are too many systematic errors in the dataset, the network will make systematic errors itself, in what is known as the "garbage in–garbage out" (GIGO) principle [136]. Some forms or sources of error can be identified relatively easily, while others—once made—are much harder to find. As in every statistical method, outliers may be present. A model trained on a small dataset is more affected by some outliers than a large dataset. This is why not only quality, but also quantity matters in machine learning. One possible solution to systematic errors is to manually remove these points from the dataset; it is also possible to use algorithms for anomaly detection, such as PCA [69,92], *t*-SNE [137,138], DBSCAN [139,140], or recurrent neural networks (LSTM networks) [111,141,142]. Recently, self-learning unsupervised neural network-based methods for anomaly detection [143] have been developed [144–146]. Next to simple outliers, there is always the possibility that the data points are actually wrong. Such data points might be one sample from an experiment in which a measurement error was made, or from a whole set of experiments that were conducted incorrectly. An example could be the results from a chemical analysis in which the apparatus was not calibrated. Training on a set of systematically false data is especially dangerous since the model will perceive the false trend as truth. Identifying such cases is possible through diligent scrutiny of the published data. This example illustrates the importance of data curation, which ensures that the data used is accurate, reliable, and reproducible.

Obviously, data can only be curated when it is available. Although decades of modeling, simulating, and experimenting have provided the chemical engineering community with a massive amount of data, this data is often stored in research laboratories or companies, and is hence not readily available. Even in a case where data is accessible, such as from an in-house database, the available data might not be completely useful for machine learning. The same applies to data extracted from research papers or patents using text-mining techniques [147]. The reason such data might not be useful is because, in general, only successful experiments are reported, while failed experiments remain unpublished [148]. Furthermore, experiments or operation conditions that seem to be nonsense to a human chemical engineer are not performed, because the engineer has insight and scientific knowledge. Machine learning algorithms, however, do not know these boundaries and not including such "trivial" data might lead to bad predictions.

5. Opportunities

The many strengths of machine learning methods present various application opportunities, and recent developments have provided ways to mitigate some of the most important criticisms. The exceptionally high execution speeds of almost any trained machine learning method makes such methods well-suited for applications in which accuracy and speed within predefined system boundaries are important. Examples of such applications include feed-forward process control and high-frequency real-time optimization [149–151]. While empirical models often lack the accuracy for these applications, detailed fundamental models are rarely fast enough to avoid computational delays. Machine learning models, trained on a fundamental model, can provide similar accuracy, yet at the computational cost of an empirical model. In this case, a model is trained on high-level data and tries to predict the difference between the empirical outcome and the true value [152,153]. Unsupervised algorithms can be used in process control applications for discovering outliers in real-time data [93]. The combination of more accurate, rapid prediction and reliable industrial data offers opportunities for the creation of digital twins and better control, leading to more efficient chemical processes.

A similar observation can be made in multiscale modeling approaches, where phenomena at a variety of different scales are modeled, resulting in a complex and strongly coupled set of equations. The potential of machine learning in such applications strongly depends on the aim of the multiscale approach. If the aim is to gain fundamental insights into the lower scale phenomena, then machine learning is not advisable, due to its black-box

nature. However, if the smaller scales are incorporated into the approach in order to obtain a more accurate model for larger scale phenomena, then machine learning could be used to replace the slow fundamental models for the smaller scales, without impacting the interpretability of the larger scale phenomena.

A final opportunity lies in providing an answer to one of the main flaws of machine learning: its non-interpretability. The issue of interpretable machine learning systems is not unique to chemical engineering problems—it is encountered in nearly any decision-making system [154–157]. An attempt has been made in the field of catalysis to rationalize what exactly machine models learn [158]. This attempt, however, still does not provide any level of direct interpretation of the model outcomes. Fig. 5 shows a workflow for explaining why a certain result is obtained. When the model outputs a good result, such as a chemical reaction predictor giving the correct product, the model should only be trusted after examining what the prediction is based on. A first step toward interpretation of the model results is to quantify the individual prediction uncertainties [159,160], as this gives an idea of the confidence the model has in its own decisions [115,161–164]. One relatively straightforward way of doing so is via ensemble modeling. This methodology has been used for decades in weather forecasting and can be used in combination with nearly any model type [165–167]. Several algorithms have also been created to determine how much certain input features influence the output [168], or to see which training points the model uses for a certain output [169,170]. When the results seem chemically or physically unreasonable, the model should be falsified instead of validated, by finding adversarial examples [159]. Furthermore, the reason is usually found in the dataset, with erroneous data or bias being present in the dataset [171,172].

Another way of making machine learning models more interpretable is to include chemically relevant and well-founded information in the models themselves. Interpretation will still require a considerable amount of postprocessing, but—if human-readable inputs are used and model architectures are not too complex—it remains a feasible task. Very complex recurrent neural networks using molecular fingerprints as input are nearly impossible to interpret, as the model input is already difficult for a human to decipher. In risk management, the "as low as reasonably practicable" (ALARP) principle is often applied [173]. Analogously, one could suggest an "as simple as reasonably possible" principle in order for machine learning models to be as interpretive as possible.

6. Threats

The accessibility of machine learning models is both a major strength and a major threat in research. While machine learning can be used by anyone with basic programming skills, it can also be misused due to a lack of algorithmic knowledge. Today, a plethora of machine learning algorithms are available, and a tremendous number of combinations of parameters and hyperparameters is possible. Even for experienced users, machine learning remains a reasoned trial-and-error method. Since researchers are often unable to explain why one algorithm works while another does not, some see machine learning as a type of modern alchemy [174]. Moreover, the majority of published articles do not provide source code, or only a pseudocode, which makes it impossible to reproduce the work [175,176]. Although chemistry and chemical engineering do not face a reproducibility crisis as much as the social sciences do [177], skepticism might grow in the community due to the increasing irreproducible use of machine learning in the field. In Gartner's hype cycle [178], machine learning and deep learning are beyond the peak of inflated expectations [179], and there is a risk of entering a period of disillusionment where interest is nearly gone. Next to irresponsible use of algorithms—and possibly more dangerous—is misinterpretation of the results. The black-box nature of the algorithms makes it difficult, and often nearly impossible, to understand why a certain result is obtained. In addition, a model might give the correct outcome for the wrong reasons [159]. Therefore, researchers should bear in mind an important rule from statistics when using machine learning: It is about the correlations, not the causations.

Another kind of unreasonable use of machine learning occurs when the model leaves the application range it is created for. The application range is determined by the training dataset and is finite. When testing unknown data points, the researcher should check whether or not these points are within the application range. When the points are outside of the range, it should be seen a warning signal for the user that the model will perform poorly [92]. The lower part of Fig. 5 depicts how the reason for obtaining a poor result is generally found by looking at the training set. Open-source applications using clustering algorithms are available for evaluating the data accuracy and its application range [180].

A last threat to applying machine learning in chemical engineering research is the growing educational gap when it comes to machine learning techniques. When applying computer and data science to chemistry and chemical engineering, it is important to understand not only the tool that is used, but also the process it is applied to. Therefore, simple training on how to use machine learning algorithms might become insufficient in the near future. Instead, a good education on AI and statistical methods will become vital in chemical engineering undergraduate programs. On the other hand, there is a need for more collaboration between computer scientists and experts on the studied topic. Whereas undertrained researchers risk a wrong use of the computational tools, computer and data scientists might obtain suboptimal results when they are not fully familiar with the topic being studied. More interdisciplinary research and a symbiosis between machine learning experts and chemical experts might be a way to avoid a phase of disillusionment.

7. Conclusions and perspectives

In the past decade, machine learning has become a new tool in the chemical engineer's toolkit. Indeed, driven by its execution speed, flexibility, and user-friendly applications, there is a strong, growing interest in machine learning among chemical engineers. On the flip side of this popularity is the risk of misusing machine learning or misinterpreting black-box results, which can potentially lead to a distrust of machine learning within the chemical engineering community. The following three recommendations can help to improve the credibility of machine learning models and turn them into an even more valuable and reliable modeling method.

First, it is important to maintain easy and open access to data and models within the community. High-quality data and open-source models encourage researchers to use machine learning as a tool and grant them the ability to focus on their topic rather than on programming and gathering data. Second, but related to the first point, is the creation of interpretable models. Since machine learning is already established in other research areas, new models for chemical applications are often inspired by existing algorithms. Therefore, the field will benefit most from studying why a certain output is generated from a given input, rather than from maintaining black boxes. The last recommendation is to invest in a profound algorithmic education. Although chemical engineers typically have very strong mathematical and modeling skills, understanding the computer science behind the graphical interface is a prerequisite for any modeler. This should also make it possible to define the

application range of the model, which is crucial for an understanding of when the model is interpolating and when it is extrapolating. This last point is definitely the most crucial: Machine learning models should be credible models, which can only be achieved by being vigilant for times when the model is being used outside of its training set.

Acknowledgements

The authors acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation (818607). Pieter P. Plehiers and Ruben Van de Vijver acknowledge financial support, respectively, from a doctoral (1150817N) and a postdoctoral (3E013419) fellowship from the Research Foundation—Flanders (FWO).

Compliance with ethics guidelines

Maarten R. Dobbelaere, Pieter P. Plehiers, Ruben Van de Vijver, Christian V. Stevens, and Kevin M. Van Geem declare that they have no conflict of interest or financial conflicts to disclose.

References

[1] Levenspiel O. Modeling in chemical engineering. Chem Eng Sci 2002;57(22–23):4691–6.
[2] Stokes GG. On the steady motion of incompressible fluids. In: Mathematical and physical papers. Cambridge: Cambridge University Press; 2009. p. 1–16. French.
[3] Navier CL. Memoire sur les lois du mouvement des fluides. Mem Acad Sci Inst Fr 1827;6:389–440. French.
[4] Prandtl L. Über flussigkeitsbewegung bei sehr kleiner reibung. In: Riegels FW, editor. Ludwig prandtl gesammelte abhandlungen. Berlin: Springer; 1904. p. 484–91. German.
[5] Siirola JJ, Powers GJ, Rudd DF. Synthesis of system designs: III. toward a process concept generator. AIChE J 1971;17(3):677–82.
[6] Venkatasubramanian V. The promise of artificial intelligence in chemical engineering: is it here, finally? AIChE J 2019;65(2):466–78.
[7] Reaxys [Internet]. Amsterdam: Elsevier; c2021 [cited 2021 Jan 4]. Available from: https://www.elsevier.com/solutions/reaxys.
[8] CAS SciFinder [Internet]. Columbus: American Chemical Society; c2021 [cited 2021 Jan 4]. Available from: https://www.cas.org/products/scifinder.
[9] ChemSpace [Internet]. Monmouth Junction: Chemspace US Inc.; c2021 [cited 2021 Jan 4]. Available from: https://chem-space.com/about.
[10] Ruddigkeit L, van Deursen R, Blum LC, Reymond JL. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. J Chem Inf Model 2012;52(11):2864–75.
[11] NIST Chemistry WebBook. Washington, DC: National Institute of Standards and Technology, US Department of Commerce. c2018 [cited 2021 Jan 4]. Available from: https://webbook.nist.gov/chemistry/.
[12] Pettit LD. The IUPAC stability constants database. Chem Int 2006;28(5):14–5.
[13] Chen G, Chen P, Hsieh CY, Lee CK, Liao B, Liao R, et al. Alchemy: a quantum chemistry dataset for benchmarking AI models. 2019. arXiv:1906.09427.
[14] Delaney JS. ESOL: estimating aqueous solubility directly from molecular structure. J Chem Inf Comput Sci 2004;44(3):1000–5.
[15] Mobley DL, Guthrie JP. FreeSolv: a database of experimental and calculated hydration free energies, with input files. J Comput Aided Mol Des 2014;28(7):711–20.
[16] Hall MA. Correlation-based feature selection for machine learning [dissertation]. Hamilton: The University of Waikato; 1999.
[17] Khalid S, Khalil T, Nasreen SA. A survey of feature selection and feature extraction techniques in machine learning. In: Proceedings of 2014 Science and Information Conference; 2014 Aug 27–29; London, UK. New York: IEEE; 2014.
[18] Xue B, Zhang M, Browne WN, Yao X. A survey on evolutionary computation approaches to feature selection. IEEE Trans Evol Comput 2016;20(4):606–26.
[19] Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. Neurocomputing 2018;300:70–9.
[20] Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection. In: Proceedings of NIPS 2013-Twenty-Seventh Annual Conference on Neural Information Processing Systems Conference. 2013 Dec 9–12; Nevada, CA, USA. New York: Neural Information Processing Systems Foundation, Inc.; 2013.
[21] Bassam A, Conde-Gutierrez RA, Castillo J, Laredo G, Hernandez JA. Direct neural network modeling for separation of linear and branched paraffins by adsorption process for gasoline octane number improvement. Fuel 2014;124:158–67.
[22] De Oliveira FM, de Carvalho LS, Teixeira LSG, Fontes CH, Lima KMG, Câmara ABF, et al. Predicting cetane index, flash point, and content sulfur of diesel–biodiesel blend using an artificial neural network model. Energy Fuels 2017;31(4):3913–20.
[23] Li H, Zhang Z, Liu Z. Application of artificial neural networks for catalysis: a review. Catalysts 2017;7(10):306.
[24] Abdul Jameel AG, Van Oudenhoven V, Emwas AH, Sarathy SM. Predicting octane number using nuclear magnetic resonance spectroscopy and artificial neural networks. Energy Fuels 2018;32(5):6309–29.
[25] Plehiers PP, Symoens SH, Amghizar I, Marin GB, Stevens CV, Van Geem KM. Artificial intelligence in steam cracking modeling: a deep learning algorithm for detailed effluent prediction. Engineering 2019;5(6):1027–40.
[26] Cavalcanti FM, Schmal M, Giudici R, Brito Alves RM. A catalyst selection method for hydrogen production through water–gas shift reaction using artificial neural networks. J Environ Manage 2019;237:585–94.
[27] Hwangbo S, Al R, Sin G. An integrated framework for plant data-driven process modeling using deep-learning with Monte-Carlo simulations. Comput Chem Eng 2020;143:107071.
[28] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 1988;28(1):31–6.
[29] Heller S, McNaught A, Stein S, Tchekhovskoi D, Pletnev I. InChI–the worldwide chemical structure identifier standard. J Cheminform 2013;5(1):7.
[30] Krenn M, Häse F, Nigam A, Friederich P, Aspuru-Guzik A. Self-referencing embedded strings (SELFIES): a 100% robust molecular string representation. Mach Learn Sci Technol 2020;1(4):045024.
[31] Amar Y, Schweidtmann AM, Deutsch P, Cao L, Lapkin A. Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. Chem Sci 2019;10(27):6697–706.
[32] Yalamanchi KK, van Oudenhoven VCO, Tutino F, Monge-Palacios M, Alshehri A, Gao X, et al. Machine learning to predict standard enthalpy of formation of hydrocarbons. J Phys Chem A 2019;123(38):8305–13.
[33] Yalamanchi KK, Monge-Palacios M, van Oudenhoven VCO, Gao X, Sarathy SM. Data science approach to estimate enthalpy of formation of cyclic hydrocarbons. J Phys Chem A 2020;124(31):6270–6.
[34] Rupp M, Tkatchenko A, Müller KR, von Lilienfeld OA. Fast and accurate modeling of molecular atomization energies with machine learning. Phys Rev Lett 2012;108(5):058301.
[35] Hansen K, Biegler F, Ramakrishnan R, Pronobis W, von Lilienfeld OA, Müller KR, et al. Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space. J Phys Chem Lett 2015;6(12):2326–31.
[36] Faber FA, Hutchison L, Huang B, Gilmer J, Schoenholz SS, Dahl GE, et al. Prediction errors of molecular machine learning models lower than hybrid DFT error. J Chem Theory Comput 2017;13(11):5255–64.
[37] Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci 2018;4(2):268–76.
[38] Liu S, Demirel MF, Liang Y. *N*-gram graph: simple unsupervised representation for graphs, with applications to molecules. In: Advances in neural information processing systems 32. 2019 Dec 8–14; Vancouver, BC, Canada. New York: Neural Information Processing Systems Foundation, Inc.; 2019.
[39] Wang S, Guo Y, Wang Y, Sun H, Huang J. SMILES-BERT: large scale unsupervised pre-training for molecular property prediction. In: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; 2019 Sep 7–10; Niagara Falls, NY, USA. New York: IEEE; 2019. p. 429–36.
[40] Chithrananda S, Grand G, Ramsundar B. ChemBERTa: large-scale self-supervised pretraining for molecular property prediction. 2020. arXiv:2010.09885.
[41] Fabian B, Edlich T, Gaspar H, Ahmed M. Molecular representation learning with language models and domain-relevant auxiliary tasks. 2020. arXiv:2011.13230.
[42] Glem RC, Bender A, Arnby CH, Carlsson L, Boyer S, Smith J. Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to ADME. IDrugs 2006;9(3):199–204.
[43] Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR modeling: where have you been? where are you going to? J Med Chem 2014;57(12):4977–5010.
[44] Sumudu PL, Steffen L. Computational methods in drug discovery. Beilstein J Org Chem 2016;12:2694–718.
[45] Unterthiner T, Mayr A, Klambauer G, Steijaert M, Wegner J, Ceulemans H, et al. Deep learning as an opportunity in virtual screening. In: Proceedings of Workshop on Machine Learning for Clinical Data Analysis, Healthcare and Genomics (NIPS2014); 2014 Dec 8–13; Montreal, QC, Canada. Linz: Johannes Kepler University Linz; 2015.
[46] Mayr A, Klambauer G, Unterthiner T, Steijaert M, Wegner JK, Ceulemans H, et al. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. Chem Sci 2018;9(24):5441–51.
[47] Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model 2019;59(8):3370–88.
[48] Duvenaud DK, Maclaurin D, Iparraguirre J, Bombarelli R, Aspure A, Adams RP, et al. Convolutional networks on graphs for learning molecular fingerprints. In: Proceedings of the 28th International Conference on Neural Information

Processing Systems; 2015 Dec 8–12; Bali, Indonesia. Cambridge: MIR Press; 2015. p. 2224–32.
[49] Coley CW, Barzilay R, Green WH, Jaakkola TS, Jensen KF. Convolutional embedding of attributed molecular graphs for physical property prediction. J Chem Inf Model 2017;57(8):1757–72.
[50] Coley CW, Jin W, Rogers L, Jamison TF, Jaakkola TS, Green WH, et al. A graph-convolutional neural network model for the prediction of chemical reactivity. Chem Sci 2018;10(2):370–7.
[51] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: Proceedings of the 30th International Conference on Neural Information Processing Systems; 2016 Dec 5–10; Barcelona, Spain. New York: Curran Associates, Inc; 2016. p. 3844–52.
[52] Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: a benchmark for molecular machine learning. Chem Sci 2017;9(2):513–30.
[53] Kearnes S, McCloskey K, Berndl M, Pande V, Riley P. Molecular graph convolutions: moving beyond fingerprints. J Comput Aided Mol Des 2016;30(8):595–608.
[54] Battaglia P, Pascanu R, Lai M, Rezende DJ, Koray K. Interaction networks for learning about objects, relations and physics. In: Proceedings of the 30th International Conference on Neural Information Processing Systems; 2016 Dec 5–10; Barcelona, Spain. New York: Curran Associates, Inc.; 2016. p. 4502–10.
[55] Schütt KT, Arbabzadah F, Chmiela S, Müller KR, Tkatchenko A. Quantum-chemical insights from deep tensor neural networks. Nat Commun 2017;8(1):13890.
[56] Jørgensen PB, Jacobsen KW, Schmidt MN. Neural message passing with edge updates for predicting properties of molecules and materials. In: Proceedings of 32nd Conference on Neural Information Processing Systems; 2018 Dec 3–8; Montreal, QC, Canada. New York: Neural Information Processing Systems Foundation, Inc.; 2018.
[57] Li Y, Tarlow D, Brockschmidt M, Zemel R. Gated graph sequence neural network. 2017. arXiv1511.05493.
[58] Winter R, Montanari F, Noé F, Clevert DA. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. Chem Sci 2018;10(6):1692–701.
[59] Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. In: Proceedings of the 34th International Conference on Machine Learning; 2017 Aug 6–11; Sydney, NSW, Australia. New York: JMLR.org; 2017. p. 1263–72.
[60] David L, Thakkar A, Mercado R, Engkvist O. Molecular representations in AI-driven drug discovery: a review and practical guide. J Cheminform 2020;12(1):56.
[61] Morgan HL. The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. J Chem Doc 1965;5(2):107–13.
[62] Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model 2010;50(5):742–54.
[63] Pattanaik L, Coley CW. Molecular representation: going long on fingerprints. Chem 2020;6(6):1204–7.
[64] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
[65] Von Lilienfeld OA. First principles view on chemical compound space: gaining rigorous atomistic control of molecular properties. Int J Quantum Chem 2013;113(12):1676–89.
[66] James CA. Daylight theory manual [internet]. Laguna Niguel: Daylight Chemical Information Systems, Inc.; c1997–2019 [cited 2021 Jan 4]. Available from: http://www.daylight.com/dayhtml/doc/theory/.
[67] Grethe G, Blanke G, Kraut H, Goodman JM. International chemical identifier for reactions (RInChI). J Cheminform 2018;10(1):22.
[68] Segler MHS, Waller MP. Neural-symbolic machine learning for retrosynthesis and reaction prediction. Chemistry 2017;23(25):5966–71.
[69] Plehiers PP, Coley CW, Gao H, Vermeire FH, Dobbelaere MR, Stevens CV, et al. Artificial intelligence for computer-aided synthesis in flow: analysis and selection of reaction components. Front Chem Eng 2020;2:5.
[70] Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: generative models for matter engineering. Science 2018;361(6400):360–5.
[71] Wei JN, Duvenaud D, Aspuru-Guzik A. Neural networks for the prediction of organic chemistry reactions. ACS Cent Sci 2016;2(10):725–32.
[72] Eyke NS, Green WH, Jensen KF. Iterative experimental design based on active machine learning reduces the experimental burden associated with reaction screening. React Chem Eng 2020;5(10):1963–72.
[73] Coley CW, Barzilay R, Jaakkola TS, Green WH, Jensen KF. Prediction of organic reaction outcomes using machine learning. ACS Cent Sci 2017;3(5):434–43.
[74] Nam J, Kim J. Linking the neural machine translation and the prediction of organic chemistry reactions. 2016. arXiv:1612.09529.
[75] Schwaller P, Gaudin T, Lányi D, Bekas C, Laino T. "Found in translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. Chem Sci 2018;9(28):6091–8.
[76] Duan H, Wang L, Zhang C, Guo L, Li J. Retrosynthesis with attention-based NMT model and chemical analysis of "wrong" predictions. RSC Adv 2020;10(3):1371–8.
[77] Lee AA, Yang Q, Sresht V, Bolgar P, Hou X, Klug-McLeod JL, et al. Molecular transformer unifies reaction prediction and retrosynthesis across pharma chemical space. Chem Commun 2019;55(81):12152–5.
[78] Schwaller P, Laino T, Gaudin T, Bolgar P, Hunter CA, Bekas C, et al. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. ACS Cent Sci 2019;5(9):1572–83.
[79] Michalski RS, Carbonell JG, Mitchell TM. A comparative review of selected methods for learning from examples. Mach Learn 2013;1:41–82.
[80] Dey A. Machine learning algorithms: a review. Int J Comput Sci Inf Technol 2016;7(3):1174–9.
[81] Pearson K. Contributions to the mathematical theory of evolution. Philos Trans R Soc Lond A 1894;185:71–110.
[82] Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol 1933;24(6):417–41.
[83] Pearson K. On lines and planes of closest fit to systems of points in space. Lond Edinb Dublin Philos Mag J Sci 1901;2(11):559–72.
[84] Van der Maaten L, Hinton G. Visualizing data using *t*-SNE. J Mach Learn Res 2008;9:2579–605.
[85] Ester M, Kriegel HP, Sander J, Xu XW. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996 Aug 2–4; Portland, OR, USA. New York: AAAI Press; 1996. p. 226–31.
[86] Palkovits R, Palkovits S. Using artificial intelligence to forecast water oxidation catalysts. ACS Catal 2019;9(9):8383–7.
[87] Likas A, Vlassis N, Verbeek J. The global *k*-means clustering algorithm. Pattern Recognit 2003;36(2):451–61.
[88] Tang J, Yan X. Neural network modeling relationship between inputs and state mapping plane obtained by FDA–*t*-SNE for visual industrial process monitoring. Appl Soft Comput 2017;60:577–90.
[89] Zheng S, Zhao J. A new unsupervised data mining method based on the stacked autoencoder for chemical process fault diagnosis. Comput Chem Eng 2020;135:106755.
[90] Gao H, Struble TJ, Coley CW, Wang Y, Green WH, Jensen KF. Using machine learning to predict suitable conditions for organic reactions. ACS Cent Sci 2018;4(11):1465–76.
[91] Vermeire FH, Green WH. Transfer learning for solvation free energies: from quantum chemistry to experiments. Chem Eng J 2020;418:129307.
[92] Pyl SP, Van Geem KM, Reyniers MF, Marin GB. Molecular reconstruction of complex hydrocarbon mixtures: an application of principal component analysis. AIChE J 2010;56(12):3174–88.
[93] Thombre M, Mdoe Z, Jäschke J. Data-driven robust optimal operation of thermal energy storage in industrial clusters. Processes 2020;8(2):194.
[94] Lee JM, Yoo C, Choi SW, Vanrolleghem PA, Lee IB. Nonlinear process monitoring using kernel principal component analysis. Chem Eng Sci 2004;59(1):223–34.
[95] Choi SW, Park JH, Lee IB. Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. Comput Chem Eng 2004;28(8):1377–87.
[96] Ning C, You F. Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. Comput Chem Eng 2018;112:190–210.
[97] Kano M, Hasebe S, Hashimoto I, Ohno H. A new multivariate statistical process monitoring method using principal component analysis. Comput Chem Eng 2001;25(7–8):1103–13.
[98] Chiang LH, Pell RJ, Seasholtz MB. Exploring process data with the use of robust outlier detection algorithms. J Process Contr 2003;13(5):437–49.
[99] Zhang X, Zou Y, Li S, Xu S. A weighted auto regressive LSTM based approach for chemical processes modeling. Neurocomputing 2019;367:64–74.
[100] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[101] Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems. Sebastopol: O'Reilly Media; 2019.
[102] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern 1991;21(3):660–74.
[103] Ho TK. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition; 1995 Aug 14–16; Montreal, QC, Canada. New York: IEEE; 1995. p. 278–82.
[104] Vapnik V. The support vector method of function estimation. In: Suykens JAK, Vandewalle J, editors. Nonlinear modeling. Boston: Springer; 1998. p. 55–85.
[105] Matsugu M, Mori K, Mitari Y, Kaneda Y. Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Netw 2003;16(5–6):555–9.
[106] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90.
[107] Shiffman D. Neural networks. In: Fry S, editor. The nature of code. Boston: Free Software Foundation; 2012. p. 444–80.
[108] Hopfield JJ. Artificial neural networks. IEEE Circuits Devices Mag 1988;4(5):3–10.
[109] Bontemps L, Cao VL, McDermott J, Le-Khac N. Collective anomaly detection based on long short-term memory recurrent neural networks. In: Proceedings of International Conference on Future Data and Security Engineering; 2016 Nov 23–25; Can Tho City, Vietnam. Cham: Springer International Publishing; 2016.
[110] Brotherton T, Johnson T. Anomaly detection for advanced military aircraft using neural networks. In: Proceedings of 2001 IEEE Aerospace Conference; 2001 Mar 10–17; Big Sky, MT, USA. New York: IEEE; 2001.

[111] Malhotra P, Vig L, Shroff G, Agarwal P. Long short term memory networks for anomaly detection in time series. In: Proceedings of 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015. 2015 Apr 22–24; Bruges, Belgium. Wallonie: i6doc; 2015.
[112] Chalapathy R, Menon AK, Chawla S. Anomaly detection using one-class neural networks. 2018. arXiv:1802.06360.
[113] Zhou S, Shen W, Zeng D, Fang M, Wei Y, Zhang Z. Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Signal Process Image Commun 2016;47:358–68.
[114] Grambow CA, Pattanaik L, Green WH. Deep learning of activation energies. J Phys Chem Lett 2020;11(8):2992–7.
[115] Scalia G, Grambow CA, Pernici B, Li YP, Green WH. Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. J Chem Inf Model 2020;60(6):2697–717.
[116] Li YP, Han K, Grambow CA, Green WH. Self-evolving machine: a continuously improving model for molecular thermochemistry. J Phys Chem A 2019;123 (10):2142–52.
[117] Grambow CA, Li YP, Green WH. Accurate thermochemistry with small data sets: a bond additivity correction and transfer learning approach. J Phys Chem A 2019;123(27):5826–35.
[118] Christensen AS, Bratholm LA, Faber FA, Anatole von Lilienfeld O. FCHL revisited: faster and more accurate quantum machine learning. J Chem Phys 2020;152(4):044107.
[119] Azlan Hussain M. Review of the applications of neural networks in chemical process control–simulation and online implementation. Artif Intell Eng 1999;13(1):55–68.
[120] Schweidtmann AM, Mitsos A. Deterministic global optimization with artificial neural networks embedded. J Optim Theory Appl 2019;180 (3):925–48.
[121] Zhu W, Sun W, Romagnoli J. Adaptive *k*-nearest-neighbor method for process monitoring. Ind Eng Chem Res 2018;57(7):2574–86.
[122] Yan S, Yan X. Using labeled autoencoder to supervise neural network combined with *k*-nearest neighbor for visual industrial process monitoring. Ind Eng Chem Res 2019;58(23):9952–8.
[123] Walker E, Kammeraad J, Goetz J, Robo MT, Tewari A, Zimmerman PM. Learning to predict reaction conditions: relationships between solvent, molecular structure, and catalyst. J Chem Inf Model 2019;59 (9):3645–54.
[124] Zahrt AF, Henle JJ, Rose BT, Wang Y, Darrow WT, Denmark SE. Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. Science 2019;363(6424):eaau5631.
[125] Han K, Jamal A, Grambow CA, Buras ZJ, Green WH. An extended group additivity method for polycyclic thermochemistry estimation. Int J Chem Kinet 2018;50(4):294–303.
[126] Settles B. From theories to queries: active learning in practice. JMLR 2011;16:1–18.
[127] Settles B. Active learning literature survey. Computer Sciences Technical Report 1648. Madison: University of Wisconsin-Madison; 2009.
[128] Clayton AD, Schweidtmann AM, Clemens G, Manson JA, Taylor CJ, Niño CG, et al. Automated self-optimisation of multi-step reaction and separation processes using machine learning. Chem Eng J 2020;384:123340.
[129] Zhang C, Amar Y, Cao L, Lapkin AA. Solvent selection for mitsunobu reaction driven by an active learning surrogate model. Org Process Res Dev 2020;24 (12):2864–73.
[130] Schütt KT, Sauceda HE, Kindermans PJ, Tkatchenko A, Müller KR. SchNet-deep learning architecture for molecules and materials. J Chem Phys 2018;148 (24):241722.
[131] Schütt KT, Kessel P, Gastegger M, Nicoli KA, Tkatchenko A, Müller KR. SchNetPack: a deep learning toolbox for atomistic systems. J Chem Theory Comput 2019;15(1):448–55.
[132] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[133] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16); 2016 Nov 2–4; Savannah, GA, USA. Northbrook: USENIX; 2016.
[134] Chollet F. Keras [internet]. San Francisco: GitHub, Inc.; 2021 Jun 18 [cited 2021 Jan 4]. Available from: https://github.com/keras-team/keras.
[135] Paszke A, Gross S, Massa F, Lerer A, Chintala S. Pytorch: an imperative style, high-performance deep learning library. In: Proceedings of 33rd Conference on Neural Information Processing Systems; 2019 Dec 8–14; Vancouver, BC, Canada. New York: Neural Information Processing Systems Foundation, Inc.; 2019.
[136] Bininda-Emonds ORP, Jones KE, Price SA, Cardilloe M, Grenyer R, Purvis A. Garbage in, garbage out. In: Bininda-Emonds ORP, editor. Phylogenetic supertrees. Berlin: Springer; 2004. p. 267–80.
[137] Schubert E, Gertz M. Intrinsic *t*-stochastic neighbor embedding for visualization and outlier detection. In: Proceedings of International Conference on Similarity Search and Applications; 2017 Oct 4–6; Munich, Germany. Berlin: Springer; 2017. p. 188–203.
[138] Perez H, Tah JHM. Improving the accuracy of convolutional neural networks by identifying and removing outlier images in datasets using *t*-SNE. Mathematics 2020;8(5):662.
[139] Çelik M, Dadaşer-Çelik F, Dokuz AŞ. Anomaly detection in temperature data using DBSCAN algorithm. In: Proceedings of 2011 International Symposium on Innovations in Intelligent Systems and Applications; 2011 Jun 15–18; Istanbul, Turkey. New York: IEEE; 2016. p. 91–5.
[140] Cassisi C, Ferro A, Giugno R, Pigola G, Pulvirenti A. Enhancing density-based clustering: parameter reduction and outlier detection. Inf Syst 2013;38 (3):317–30.
[141] Fernando T, Denman S, Sridharan S, Fookes C. Soft + hardwired attention: an LSTM framework for human trajectory prediction and abnormal event detection. Neural Netw 2018;108:466–78.
[142] Filonov P, Lavrentyev A, Vorontsov A. Multivariate industrial time series with cyber-attack simulation: fault detection using an LSTM-based predictive data model. 2016. arXiv:1612.06676.
[143] Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv 2009;41(3):1–58.
[144] Ahmad S, Lavin A, Purdy S, Agha Z. Unsupervised real-time anomaly detection for streaming data. Neurocomputing 2017;262:134–47.
[145] Amini M, Jalili R, Shahriari HR. RT-UNNID: a practical solution to real-time network-based intrusion detection using unsupervised neural networks. Comput Secur 2006;25(6):459–68.
[146] Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Proceedings of International Conference on Information Processing in Medical Imaging 2017; 2017 Jun 25–30; Boone, KY, USA. Cham: Springer International Publishing; 2017. p. 146–57.
[147] Schneider N, Lowe DM, Sayle RA, Tarselli MA, Landrum GA. Big data from pharmaceutical patents: a computational analysis of medicinal chemists' bread and butter. J Med Chem 2016;59(9):4385–402.
[148] Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, et al. Machine-learning-assisted materials discovery using failed experiments. Nature 2016;533(7601):73–6.
[149] Wu Z, Rincon D, Christofides PD. Real-time adaptive machine-learning-based predictive control of nonlinear processes. Ind Eng Chem Res 2020;59 (6):2275–90.
[150] Zhang Z, Wu Z, Rincon D, Christofides P. Real-time optimization and control of nonlinear processes using machine learning. Mathematics 2019;7(10):890.
[151] Powell BKM, Machalek D, Quah T. Real-time optimization using reinforcement learning. Comput Chem Eng 2020;143:107077.
[152] Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA. Big data meets quantum chemistry approximations: the Δ-machine learning approach. J Chem Theory Comput 2015;11(5):2087–96.
[153] Bikmukhametov T, Jäschke J. Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. Comput Chem Eng 2020;138:106834.
[154] Gunning D, Aha DW. DARPA's explainable artificial intelligence program. AI Mag 2019;40(2):44–58.
[155] Abdul A, Vermeulen J, Wang D, Lim BY, Kankanhali M. Trends and trajectories for explainable, accountable and intelligible systems: an HCI research agenda. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; 2018 Apr 21; Montreal, QC, Canada. New York: Association for Computing Machinery; 2018.
[156] Lepri B, Oliver N, Letouzé E, Pentland A, Vinck P. Fair, transparent, and accountable algorithmic decision-making processes. Philos Technol 2018;31 (4):611–27.
[157] Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. Sci Robot 2017;2(6):eaan6080.
[158] Kammeraad JA, Goetz J, Walker EA, Tewari A, Zimmerman PM. What does the machine learn? Knowledge representations of chemical reactivity. J Chem Inf Model 2020;60(3):1290–301.
[159] Kovács DP, McCorkindale W, Lee AA. Quantitative interpretation explains machine learning models for chemical reaction prediction and uncovers bias. Nat Comm 2021;12:1695.
[160] Preuer K, Klambauer G, Rippmann F, Hochreiter S, Unterthiner T. Interpretable deep learning in drug discovery. In: Samek W, Montavon G, Vedaldi A, Hansen LK, Müller KR, editors. Explainable AI: interpreting, explaining and visualizing deep learning. Cham: Springer International Publishing; 2019. p. 331–45.
[161] Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. Nat Mach Intell 2019;1 (1):20–3.
[162] Mohamed L, Christie MA, Demyanov V. Comparison of stochastic sampling algorithms for uncertainty quantification. SPE J 2010;15(01):31–8.
[163] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. 2016. arXiv:1506.02142v6.
[164] Fridlyand A, Johnson MS, Goldsborough SS, West RH, McNenly MJ, Mehl M, et al. The role of correlations in uncertainty quantification of transportation relevant fuel models. Combust Flame 2017;180:239–49.
[165] Parker WS. Ensemble modeling, uncertainty and robust predictions. Wiley Interdiscip Rev Clim Change 2013;4(3):213–23.
[166] Gneiting T, Raftery AE. Weather forecasting with ensemble methods. Science 2005;310(5746):248–9.
[167] Derome J. On the average errors of an ensemble of forecasts. Atmos Ocean 1981;19(2):103–27.
[168] Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. 2017. arXiv:1703.01365.
[169] Datta A, Sen S, Zick Y. Algorithmic transparency via quantitative input influence: theory and experiments with learning systems. In: Proceedings of

2016 IEEE Symposium on Security and Privacy (SP); 2016 May 22–26; San Jose, CA, USA. New York: IEEE. p. 598–617.
[170] Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM Comput Surv 2018;51(5):93.
[171] Lin AI, Madzhidov TI, Klimchuk O, Nugmanov RI, Antipin IS, Varnek A. Automatized assessment of protective group reactivity: a step toward big reaction data analysis. J Chem Inf Model 2016;56(11):2140–8.
[172] Pesciullesi G, Schwaller P, Laino T, Reymond JL. Transfer learning enables the molecular transformer to predict regio- and stereoselective reactions on carbohydrates. Nat Commun 2020;11(1):4874.
[173] Melchers RE. On the ALARP approach to risk management. Reliab Eng Syst Saf 2001;71(2):201–8.
[174] Hutson M. Has artificial intelligence become alchemy? Science 2018;360(6388):478.
[175] Hutson M. Artificial intelligence faces reproducibility crisis. Science 2018;359(6377):725–6.
[176] Gundersen OE, Kjensmo S. State of the art: reproducibility in artificial intelligence. In: Proceedings of The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18); 2018 Feb 2–7; New Orleans, LA, USA. Palo Alto: AAAI Press; 2018. p. 1644–51.
[177] Baker M. 1,500 scientists lift the lid on reproducibility. Nature 2016;533:452–4.
[178] Fenn J, Linden A. Understanding Gartner's hype cycles. Report. Stamford: Gartner, Inc.; 2003 May. Report No: R-20-1971.
[179] Sicular S. Vashisth S. Hype cycle for artificial intelligence, 2020 [Internet]. Reading: CloudFactory; 2020 Jul 27 [cited 2021 Jan 4]. Available from: https://www.cloudfactory.com/reports/gartner-hype-cycle-for-artificial-intelligence.
[180] Symoens SH, Aravindakshan SU, Vermeire FH, De Ras K, Djokic MR, Marin GB, et al. QUANTIS: data quality assessment tool by clustering analysis. Int J Chem Kinet 2019;51(11):872–85.