Article:
Yan, Y., Borhani, T.N., Subraveti, S.G. et al. (16 more authors) (2021) Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS) – a state-of-the-art review. Energy and Environmental Science, 14 (12). pp. 6122-6157. ISSN 1754-5692

https://doi.org/10.1039/d1ee02395k

Energy & Environmental Science

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal | View Issue

Cite this: *Energy Environ. Sci.*, 2021, **14**, 6122

Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS) – a state-of-the-art review

Yongliang Yan, *[ab] Tohid N. Borhani, [c] Sai Gokul Subraveti, [d] Kasturi Nagesh Pai, [d] Vinay Prasad, [d] Arvind Rajendran, [d] Paula Nkulikiyinka, [a] Jude Odianosen Asibor, [a] Zhien Zhang, [e] Ding Shao, [f] Lijuan Wang, [g] Wenbiao Zhang, [f] Yong Yan, [g] William Ampomah, [h] Junyu You, [hi] Meihong Wang, [j] Edward J. Anthony, [a] Vasilije Manovic [a] and Peter T. Clough *[a]

Received 3rd August 2021,
Accepted 1st November 2021

DOI: 10.1039/d1ee02395k

rsc.li/ees

Carbon capture, utilisation and storage (CCUS) will play a critical role in future decarbonisation efforts to meet the Paris Agreement targets and mitigate the worst effects of climate change. Whilst there are many well developed CCUS technologies there is the potential for improvement that can encourage CCUS deployment. A time and cost-efficient way of advancing CCUS is through the application of machine learning (ML). ML is a collective term for high-level statistical tools and algorithms that can be used to classify, predict, optimise, and cluster data. Within this review we address the main steps of the CCUS value chain (CO_2 capture, transport, utilisation, storage) and explore how ML is playing a leading role in expanding the knowledge across all fields of CCUS. We finish with a set of recommendations for further work and research that will develop the role that ML plays in CCUS and enable greater deployment of the technologies.

Broader context

Carbon capture, utilisation and storage (CCUS) is well recognised to play a critical role in future decarbonisation efforts to meet Paris Agreement goals and net zero emissions targets. Machine learning (ML) is a collective term for high-level statistical tools and algorithms that can be used to classify, predict, optimise, and cluster data. ML has been applied to CCUS technologies as a powerful tool to accelerate their development. This work presents a state-of-the-art review of ML applications in CO_2 capture, transport, storage, and utilisation, and provides perspectives for the field. In this manuscript, the authors provide a set of recommendations for further work and research that will help develop the role that ML plays in CCUS and enable greater deployment of CCUS technologies.

1. Introduction

As atmospheric CO_2 concentrations surpass yet another milestone (>420 ppm in April 2021[1]), climate change continues to be described as the biggest threat to humanity and global security.[2] It is for this reason that global efforts to decarbonise all sectors of society through Nationally Determined Contributions (NDCs) have begun to be strengthened and provides the backdrop for the COP26 discussions.[3]

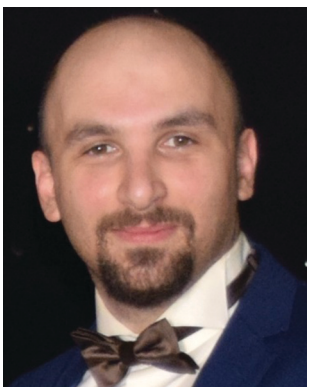
The recent COVID-19 pandemic has provided the opportunity to foresee a 'new normal' where lifestyles can be radically different, and a sense of national contribution can be understood.

[a] *Energy and Power Theme, Cranfield University, Bedfordshire, MK43 0AL, UK. E-mail: yongliang.yan@newcastle.ac.uk, p.t.clough@cranfield.ac.uk*
[b] *Materials, Concept and Reaction Engineering (MatCoRE) Group, School of Engineering, Newcastle University, Merz Court, Newcastle Upon Tyne, NE1 7RU, UK*
[c] *School of Engineering, Division of Chemical Engineering, University of Wolverhampton, Wolverhampton, WV1 1LY, UK*
[d] *Department of Chemical and Materials Engineering, Donadeo Innovation Centre for Engineering, University of Alberta, 9211-116 Street NW, Edmonton, Alberta, T6G 1H9, Canada*
[e] *Department of Chemical and Biomedical Engineering, West Virginia University, Morgantown, WV, 26506, USA*
[f] *School of Control and Computer Engineering, North China Electric Power University, Beijing, 102206, P. R. China*
[g] *School of Engineering, University of Kent, Canterbury, Kent, CT2 7NT, UK*
[h] *Petroleum Recovery Research Centre, New Mexico Tech, Socorro NM, 87801, USA*
[i] *School of Petroleum and Natural Gas Engineering, Chongqing University of Science and Technology, Chongqing, 401331, China*
[j] *Department of Chemical and Biological Engineering, University of Sheffield, Sheffield S1 3JD, UK*

Furthermore, the COVID-19 pandemic has led to governments around the world utilising this change as an opportunity to "Build Back Better" with "Green Growth" and a "Green Industrial Revolution".[4–8] Part of these recovery plans involve the deployment of CCUS at significant scales in the coming decades to meet net zero pledges and limit warming to 1.5 °C. CCUS is absolutely crucial for the decarbonisation of many sectors that cannot be decarbonised by other process changes (*e.g.*, cement, iron and steel). The roll out of Carbon Capture and Storage (CCS) is planned to achieve 10 Mt CO_2 captured per year by 2030 in the UK, with other similar commitments globally.[9] In addition, all negative emissions technologies (NET), such as direct air capture (DAC) and Biomass Energy with Carbon Capture and Storage (BECCS) technologies require the deployment of CCUS. These technologies allow otherwise stranded fossil fuel in the power sector to continue to be used at a much higher level and reduces the abatement requirements of fossil fuels (including natural gas) to a 28–33% level, instead of a 46–57% level while staying below a 2 °C temperature target.[10] Moreover, there is also a growing awareness in the EU and countries like Canada that meeting net zero emissions by 2050[11] and 2060 for China,[12] unconventional methods such as DAC will be required.[13] A similar view is developing in the USA, that negative emissions technologies are required to meet current climate goals by 2050 and without them, the US net zero initiative will fail.[14] Moreover, the idea that a 100% wind, water and solar scenarios are even achievable by 2050 has also received challenges.[15] In light of this, more affordable CCUS, is not just desirable, but also essential. However, a general review of CCUS technology and its roll out is available from others,

Yongliang Yan

Yongliang Yan is a Research Associate in the MatCoRE group at Newcastle University. He obtained his PhD in Energy and Power from Cranfield University, where he worked on the application of high-temperature solid looping cycles for CO_2 capture, hydrogen production and thermochemical energy storage. His current research is focused on scaling up a novel chemical-looping concept for low-carbon hydrogen production, and applying machine learning and AI in materials development, process simulation and optimisation.



Tohid N. Borhani

Tohid N. Borhani is currently Lecturer in Chemical Engineering at University of Wolverhampton. He was Assistant Professor in Chemical Engineering at Heriot-Watt University. He also was Research Associate at Cranfield University, The University of Sheffield, and Imperial College London for several years. He has published more than 40 journal papers in high ranked journals. His main research area is process modelling and simulation (with a focus on carbon capture); data-driven chemometric modelling, and computational chemistry; computational fluid dynamics (CFD) of reactor networks; application of artificial intelligence in chemical engineering; and energy conversion and storage.

Sai Gokul Subraveti

Sai Gokul Subraveti received his BTech (2014) from NIT Warangal, India. He then moved to Canada where obtained his MSc (2017) and PhD (2021) degrees in Chemical Engineering from the University of Alberta. He carried out his doctoral studies under the supervision of Prof. Arvind Rajendran, Prof. Vinay Prasad, and Prof. Zukui Li. His research interests focus on process design and optimization of adsorption processes, machine learning, and techno-economic assessments, primarily for CO_2 capture. He will be joining as a Research Scientist at SINTEF Energy Research, Norway.

Yong Yan

Yong Yan received his BEng and MSc degrees in instrumentation and control engineering from Tsinghua University, China, in 1985 and 1988, respectively, and PhD degree in flow measurement from University of Teesside, UK, in 1992. He is Professor of Electronic Instrumentation and Director of Innovation at School of Engineering, University of Kent, UK. In recognition of his contributions to pulverized fuel flow metering and burner flame imaging, he was elected as an IEEE Fellow in 2011 and a Fellow of the Royal Academy of Engineering in 2020. His research interests include multiphase flow measurement, combustion instrumentation, and intelligent condition monitoring.

so the authors will not go into details, explaining the basic mechanics of CCUS processes.[16]

The use of machine learning (ML) has increased for a multitude of applications due to the growth in computing power in recent years, this is true for CCUS applications as well. ML offers the potential to identify links between data/results that aren't readily identifiable, and it also provides alternative lower computing cost pathways. Within the field of CCUS, ML has begun to be utilised to evaluate new CO_2 sorbents and oxygen carrier materials,[17] simulate, control and operate capture processes,[18–23] simplify process economics, predict CO_2 solubilities in solvents and CO_2 capture capacities in adsorbents,[24–26] improve the accuracy of multiphase flowmeters used for CO_2 pipelines,[27] and predict leaks from CO_2 wells;[28] each with the aim of advancing the field of CCUS in a cost and time effective manner. Meanwhile, it is also worth noting that ML is data-driven technology, and its performance usually depends on the size and quality of database. In some areas of CCUS, the available data size can be limited to only a few dozens of datapoints and some of the raw data may not even be published openly, which will limit researchers in applying ML in those areas. Moreover, ML is a powerful tool for complex and nonlinear problems. It may not be suitable for applications that can be easily solved by numerical methods. Another big challenge for ML is it is difficult to extract the new knowledge from ML models to form general conclusions and scientific laws. Researchers in CCUS should consider what new information they can extract from ML models before applying ML in their research. Nevertheless, ML in CCUS is still relatively new and there is much yet to be studied.

Past studies in ML in CCUS are scattered within the literature and there has been no previous attempt to reconcile this information, gathered along the entire CO_2 supply chain, systemically into a critical review and summary and set out a clear pathway forward. A detailed and systematic critical analysis of previous research will lead to an acceleration of CCUS commercialisation and an expansion of ML in all areas of CCUS, this forms the main motivation behind this review.

2. Machine learning algorithms

ML is a subset of artificial intelligence (AI) that involves the study of computer algorithms that allow computer programs to automatically improve through experience.[29,30] Its advantages include ease of trends and pattern identification, minimal human intervention (automation), ability to improve continuously, as well as high efficiency in the handling of multi-dimensional and multi-variety data.[29,31] Its application is however sometimes limited by factors such as ethics, lack of physical constraint, data availability and quality, misapplication as well as interpretability.[32]

The dependence of ML modelling on data presents some challenges in terms of availability, quantity as well as quality. Given this dependence, if the sourced data contains human biases and prejudices, then the decision of models developed from such data may inherit such biases, consequently leading to unfair and wrong decisions. Closely associated with the aspect of data is the challenge of dimensionality (the curse of dimensionality). This refers to all the problems that arise when working with data in higher dimensions (large number of data features) that did not exist in lower dimensions.[33] This leads to overfitting resulting in poor performance of the model. In order to avoid this, dimensionality reduction, which is the transformation of high-dimensional data into a meaningful representation of reduced dimensionality is carried out.[34] This data preprocessing improves the performance of the data, reduces training time and computational resources as well as noise removal.[35] Dimensionality reduction methods include: Principal

William Ampomah

William Ampomah is Assistant Professor of Petroleum Engineering at New Mexico Tech (NMT), USA. Dr Ampomah is also the Section Head of the REACT group at the Petroleum Recovery Research Center (PRRC) at NMT. He is a Lecturer at KNUST, Ghana. He is Principal Investigator and/or Co-Principal Investigator on at least five (5) US Department of Energy grants in research areas such as Enhanced oil recovery, CO_2 Storage, subsurface geomechanics, subsurface monitoring, rare earth elements. Dr Ampomah has published over 50 papers in areas of enhanced oil recovery, CO_2 storage, reservoir characterization, application of machine learning in numerical simulation and optimization.

Peter T. Clough

Peter Clough is a Senior Lecturer in Energy Engineering at Cranfield University. Dr Clough obtained an MEng in Environmental Engineering from The University of Nottingham and a PhD in Chemical Engineering at Imperial College London. In 2017 Dr Clough joined Cranfield University's Energy and Power Theme to continue his work into hydrogen and decarbonisation technologies. His research interests include hydrogen production by sorption enhanced steam methane reforming and cheminformatic material development for use in thermochemical processes.

Component Analysis (PCA), Factor Analysis, Linear Discriminant Analysis (LDA), Multi-dimensional Scaling (MDS), Isometric Feature Mapping (Isomap), t-distributed Stochastic Neighbour Embedding (t-SNE) and auto-encoders.[33,34]

ML model interpretation is another major challenge of deploying ML. This is as a result of the black-box nature of many ML models in which humans are unable to explain the decision-making logic of the ML model despite obtaining high predictive accuracy. This crucial weakness impacts not only on ethics but also on accountability, trust, transparency, safety and industrial liability.[36] To address this limitation and given the importance of openness in scientific research, several approaches have been reported with some even deployed at the cost of sacrificing accuracy. Some of these methods and techniques include; decision tree, feature importance, sensitivity analysis, partial dependence plots, activation maximization, explainable neural network (XNN), local interpretable model-agnostic explanation (LIME), shapley additive exPlanations (SHAP), Deep Learning Important FeaTures (DeepLIFT) explanation method and Treeinterpreter.[36,37] Key factors to consider in building interpretable ML models have also been reported to include but not be limited to the degree of white-box modelling, data visualisation, usability, model visualisation, variable importance, accuracy, fairness, and sensitivity residuality.[36,38] In the application of ML to CCUS, it is recommended to aim for the use and development of interpretable models with competitive levels of predictive accuracy.

Fig. 1 presents the types of ML and respective areas of application. There are three main types of ML: supervised, unsupervised and reinforcement learning. The supervised ML, which is the most commonly used of the three is usually applied when the input–output data is known. It involves training the ML models to learn the relationship between the given inputs and associated output values.[39] If the available dataset consists of only input values (no labels), unsupervised ML can be used in an attempt to identify trends, structure, patterns or clustering in the input data.[40] Reinforcement learning is a ML technique that enables an agent to learn in an interactive environment by trial and error using feedback from its actions and experiences.[41] The execution of any of the types of ML can be done through the application of the appropriate algorithm. A brief description of common ML algorithms is presented in Table 1.

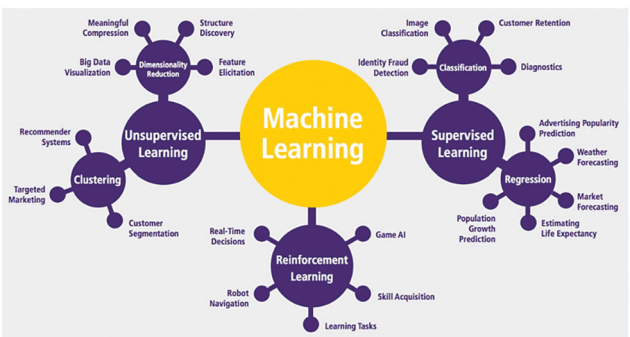


Fig. 1 Types and applications of ML.[42]

Other ML algorithms include *K*-nearest neighbour, density-based spatial clustering of applications with noise (DBSCAN), recommender systems, genetic algorithm, gradient boosting trees and particle swarm algorithms. Given the numerous types of ML models, the choice of model to be deployed in a particular application is very much dependent on factors such as task type, type and structure of expected output, type and size of data, accuracy-interpretability consideration, number of data features, linearity, available computational time as well as model complexity.[39] It is important to note that in many applications, multiple algorithms are usually combined (referred to as ensemble algorithms) to improve model performance accuracy and robustness. Information and learning resources on ML are readily available and accessible on various websites and online platforms. Table 2 presents some publicly accessible tools and resources for general purpose ML and CCUS related application.

3. Machine learning in CO_2 capture

3.1 Machine learning in CO_2 absorption

ML has wide application in modelling and analysis of different separation units such as distillation, absorption, and regeneration columns.[43] This section will focus on the research that has been done in the past decade to model and analyse different aspects of CO_2 absorption process using different solvents. It includes process modelling, simulation, and optimisation; thermodynamic analysis; and solvents selection and design. These four main areas of application of ML in CO_2 absorption are discussed in this section. Selected studies and research related to each part are also reviewed and discussed.

3.1.1 Process simulation and optimisation

3.1.1.1 Background and challenges of mathematical and optimisation models. Due to the complex governing phenomena in absorption (especially chemical absorption, which includes mass transfer and chemical reactions) modelling and simulation of solvent-based carbon capture is a time consuming and intensive job. Two common approaches to model CO_2 absorption process are equilibrium-stage model and non-equilibrium stage models. The set of equations that describe the equilibrium-stage model for the separation processes are termed the MESH equations (*i.e.*, the mass balance equations, equilibrium relations, summation relations and the enthalpy equations). In the case of non-equilibrium stage models, the separation processes are described by the MERQ equations (*i.e.*, the material balance equations, energy balance equations, rate (transfer rate) equations and the equilibrium relations).[44] In addition to MESH and MERQ equations, numerous parameters related to physical properties and transport properties such as density, viscosity, thermal conductivity, heat capacity, diffusivity coefficient, and mass and heat transfer coefficients must be considered in the model. The mass and heat balances must be considered for both liquid and gas phases and complex mathematic methods must be applied to solve the obtained set of algebraic and differential equations.

Table 1 Common ML algorithms[42,54]

Algorithm name	Task type	Description
Linear regression	Regression	By fitting a linear model with coefficients, this algorithm correlates each data feature to the output, thus assisting in predicting future values
Logistic regression	Classification	A classification algorithm that predicts the likelihood of a dependent variable (usually binary) belonging to a category
Decision tree	Regression and classification	This interpretable algorithm performs by splitting values of data features into branches at decision nodes until a final decision output is established
Naïve Bayes	Regression and classification	This algorithm is based on the Bayes' theorem which updates the prior knowledge of an event with the independent probability of each feature that can affect the event
Support Vector Machines (SVMs)	Regression, classification, and outlier detection	This algorithm operates by transforming the required data and determining the optimal boundary (hyperplane) between the various outputs
Random forest	Regression and classification	The algorithm is an ensemble of decision trees characterised by improved accuracy. It operates by generating a multitude of decision trees and uses either the modal vote or average prediction for classification or regression tasks respectively
Artificial Neural Network (ANN)	Regression, classification, and clustering	This algorithm which is modelled after the biological neurons of the brain consists of several layers with interconnected artificial neurons performing various data transformations to obtain the required output
K-means clustering	Clustering	This centroid-based algorithm clusters unlabelled data points by their similarity of characteristics determined by the model without human interference
Hierarchical clustering	Clustering	This algorithm splits clusters along a hierarchical tree to form a classification system
Gaussian mixture model	Clustering	This unsupervised algorithm clusters data by estimating the density distribution of the dataset
AdaBoost	Regression and classification	This is an ensemble algorithm that combines multiple weak algorithms to obtain an improved output
Principal component analysis (PCA)	Dimension reduction	This algorithm is often used to reduce the dimensionality of large data sets without distorting its characteristics (though it is not strictly a ML algorithm in its own right)

Table 2 Publicly accessible learning resources and tools related to ML

Name	Description	URL
General-purpose machine-learning frameworks[40]		
Caret	Package for ML in R	https://topepo.github.io/caret
Deeplearning4j	Distributed deep learning for Java	https://deeplearning4j.org
H2O.ai	Machine-learning platform written in Java that can be imported as a Python or R library	https://h2o.ai
Keras	High-level neural-network API written in Python	https://keras.io
Mlpack	Scalable machine-learning library written in C++	https://mlpack.org
Scikit-learn	Machine-learning and data-mining member of the scikit family of toolboxes built around the SciPy Python library	https://scikit-learn.org
Weka	Collection of machine-learning algorithms and tasks written in Java	https://cs.waikato.ac.nz/ml/weka
TensorFlow	An open source for numerical and large-scale ML	https://www.tensorflow.org
ML tools for CCUS		
COMBO	Python library with emphasis on scalability and efficiency	https://github.com/tsudalab/combo
DeepChem	Python library for deep learning of chemical systems	https://deepchem.io
MatMiner	Python library for assisting ML in materials science	https://hackingmaterials.github.io/matminer
NOMAD	Collection of tools to explore correlations in materials datasets	https://analytics-toolkit.nomad-coe.eu
Silicone v1.0.0	An open-source Python package for inferring missing emissions data for climate change research	https://github.com/GranthamImperial/silicone
Carboncalc	Tools to calculate growth statistics for individual urban trees such as for estimating carbon storage	https://github.com/adhollander/carboncalc
Fair	Python package that takes emissions of greenhouse gases, aerosol and ozone precursors, and converts these into greenhouse gas concentrations, radiative forcing and temperature change	https://pypi.org/project/fair
pyGAPS	A Python framework for adsorption data analysis and isotherm fitting	https://github.com/pauliacomi/pyGAPS

As many of the models used to predict the physical properties are experimental based models, there is considerable error and deviation in the prediction of different parameters that directly affect the results of the process model.[45] It should be noted that in the case of dynamic simulation which contains partial differential equations (PDE), the initial points to solve the problem is a critical aspect of the modelling job. Finding these can be a very tedious and time-consuming process.

Despite all these above-mentioned weaknesses and drawbacks, applying ML to model and optimise the solvent-based carbon capture is attracting increasing attention. Methods like ANN, adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS), support vector regression (SVR), radial basis function (RBF), and genetic programming (GP) can examine complex interaction between inputs to the model and predict the target (usually CO_2 capture levels and rate of absorption of CO_2). It should be noted that as experimental process data acquisition is frequently inadequate for various types of solvents, the majority of the researchers first developed a first principle mathematical model in a process simulator (such as Aspen Plus®, Aspen HYSYS®, and gPROMS®) and collected the data from that model. Then the collected data are used to develop the ML-based model. The ML-based models can predict the required targets with acceptable accuracy and be used easily for future studies.[46,47]

3.1.1.2 Review of the ML-based process modelling and optimisation studies. Sipöcz *et al.*[46] used a multilayer feed-forward neural network to capture and model the non-linear relationship between inputs and outputs of the solvent-based CO_2 capture process. The data used for training and validation of the ANN were obtained using the process simulator CO2SIM. The trained model was then used for finding the optimum operation for the example plant with respect to the lowest possible specific steam duty and maximum CO_2 capture rate. The authors reported that the average value of the errors for the prediction of specific reboiler duty was less than 0.2% and the maximum error was 3.1%. The prediction of solvent rich loading and amount CO_2 captured had a maximum error lower than 2.8% and 0.17% respectively.

Nuchitprasittichai and Cremaschi[48] used response surface methodology (RSM) and ANN to minimise the capture cost of CO_2 using different amines. RSM uses local searches to estimate an appropriate direction to reduce the objective function while ANN uses simulation to build a global surrogate model of the objective function over the entire decision space and solves the optimization problem using a global solver.

The structure of the algorithm in this study is presented in Fig. 2. The first step of the algorithm is the determination of the appropriate sample size to construct the ANN, the second step is optimization by using the constructed ANN with the sample size obtained from the first step as the objective function. The results showed that the number of simulations, the minimum CO_2 capture cost, and the percent error, for both methods were close to each other. The data required for the study was provided from an Aspen HYSYS® simulation.

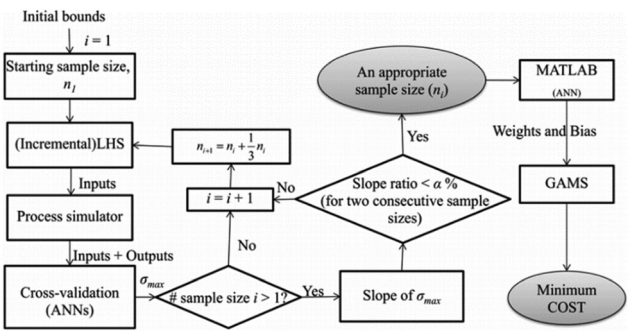


Fig. 2 Structure of the algorithm to perform optimisation.[48]

Li *et al.*[49] considered different parameters namely inlet flue gas flow rate, CO_2 concentration in inlet flue gas, the pressure of the flue gas, the temperature of the flue gas, lean solvent flow rate, monoethanolamine (MEA) concentration and the temperature of lean solvent as input to predict the CO_2 capture rate and CO_2 capture level using bootstrap aggregated neural networks. The required data to develop ML models were extracted from first principle steady-state and dynamic models developed in gPROMS®. It should be noted that both absorber and stripper were included in their model. Zhan *et al.*[50] studied the simultaneous absorption of CO_2 and H_2S in a mixture of *N*-methyl diethanolamine (MDEA) and piperazine (PZ) in a rotating packed bed (RPB) experimentally. The authors developed an ANN model to predict the absorption efficiencies of H_2S and CO_2 and mass-transfer coefficient (K_Ga).

Shalaby *et al.*[51] considered a fine tree, Matern Gaussian Process Regression (GPR), rational quadratic GPR, squared exponential GPR and feed-forward ANN models to predict the different output from CO_2 capture unit using MEA solution. Reboiler duty, condenser duty, reboiler pressure, flow rate, temperature, and the pressure of the flue gas were considered as inputs to the models and the system energy requirements, capture rate, and the purity of condenser outlet stream were the output of the models. The required data were obtained from the gPROMS process builder and the results of the models indicated high prediction accuracy.

After the development of the models, the authors developed a non-linear programming (NLP) problem and solved it using sequential quadratic programming algorithm (SQP) and genetic algorithm optimization on the surrogate model to determine the optimal operating conditions. This study showed that ML-based methods could be used to model and optimise the CO_2 capture unit appropriately. Wu *et al.*[23] developed an intelligent predictive controller (IPC) for a large-scale solvent-based post-combustion CO_2 capture process, and an ANN model was trained to predict the dynamics of the CO_2 capture process. The results indicated that the IPC demonstrated fast control of the CO_2 capture level and reduced the fluctuations in re-boiler's temperature significantly.

3.1.2 Thermodynamic analysis

3.1.2.1 Background of mathematical thermodynamic analysis. Thermodynamic analysis for solvent-based carbon capture can be classified in terms of two main tasks. One of them is chemical equilibrium calculation and the other is physical equilibrium calculation. Chemical equilibrium (speciation equilibrium) calculations provide the concentrations of different species in a solution. The modelling of speciation equilibrium is used in the calculation of enhancement factor, vapour–liquid equilibrium (VLE) modelling, and calculation of the CO_2 loading value. Implementation of chemical equilibrium calculation requires extensive knowledge about the chemical reactions in

the system and all the related parameters and models for kinetic reactions and equilibrium constants for equilibrium reactions in the combination of mass transfer balances.[52]

On the other hand, VLE modelling for the CO_2 capture system is a challenging task because of the non-ideal nature of the liquid phase (due to the existence of different types of interactions between ions and molecules), lack of accurate model parameters as well as the availability and quality of solubility data. In addition, an equation of state (EOS) such as Peng–Robinson, SAFT, and Soave–Redlich–Kwong is necessary. Furthermore, an activity coefficient-based model for instance Electrolyte NRTL, Wilson, and Extended UNIQUAC is also required to do the VLE calculations. The programming and implementation of these thermodynamic models, EOS and activity coefficients models is a complex and time-consuming job.[44]

3.1.2.2 Review of the ML-based thermodynamic modelling studies. As mentioned, thermodynamic modelling and calculation of solvent-based carbon capture is a tedious task. There are many studies in recent years where researchers used ML methods to perform thermodynamic analysis of CO_2 capture in different types of solvents and these will be discussed below.

Baghban *et al.*[53] compared the predictive capability of four ML models to evaluate the CO_2 solubility in 67 ionic liquids (ILs). They used the Least Square Support Vector Machine (LSSVM), ANFIS, Multi-Layer Perceptron Artificial Neural Network (MLP-ANN), and Radial Basis Function Artificial Neural Network (RBF-ANN). The solubility is considered as a function of different parameters such as operational temperature, pressure accompanied with the properties of ILs including the critical temperature, critical pressure and, acentric factor (ω). LSSVM model showed the best statistical performance in comparison to other methods.

Ghiasi and Mohammadi[55] used a Classification and Regression Tree (CART) method in modelling CO_2 solubility in different ILs as a function of system's temperature and pressure and properties of ILs including critical temperature, critical pressure, and acentric factor. A tree-based model was developed using 5330 experimental data points of CO_2 solubility in 66 different ILs. Findings reveal that the proposed model's outcomes are in excellent agreement with the corresponding experimental values. The presented model shows an average absolute relative deviation equal to 0.04% and provides considerably better estimations than the previously published ML based models.

Garg *et al.*[56] studied the CO_2 solubility in aqueous sodium salt of L-phenylalanine (Na-Phe) for different concentrations, temperatures and CO_2 pressure range, experimentally. Kent–Eisenberg and ANN models were used to model and correlate the solubility data. ANN showed better results in comparison to Kent–Eisenberg thermodynamic models.

Li *et al.*[57] compared several thermodynamic models (Kent–Eisenberg,[52] Austgen,[58] Hu–Chakma,[59] Liu *et al.*[60]) with two types of ANN models (back-propagation neural network (BPNN) and (RBF-NN)) to predict the CO_2 solubility in 3-dimethylamino-1-propanol (3DMA1P) solution for different operating conditions. The authors reported that absolute average deviation (ADD) of thermodynamic models were almost three times more than the ADD of ANN models. Babamohammadi *et al.*[61] presented experimental data of VLE for CO_2 absorption in the mixture of MEA and glycerol and then used these data to develop the ANN model to predict the VLE data. Yarveicy *et al.*[62] presented an extra trees model to predict the CO_2 loading in different chemical solvents using solubility data from the literature. The results of the extra trees model were compared to LSSVM, MLP-ANN, ANFIS, and RBF-ANN models in the literature. The authors reported a coefficient of determination (R^2) of 0.9993 and an average absolute relative deviation in percent (AARD%) of 0.15 for this model. Soroush *et al.*[63] applied ANFIS to develop a precise temperature-dependent ML model to correlate the CO_2 loading of amino acid salt solutions for different types of amino acids. This model was used to perform sensitivity analysis as well.

3.1.3 Prediction of properties

3.1.3.1 Background and challenges of developing property models. The models developed to predict the different types of properties could be empirical, semi-empirical, and theoretical. The objective is making a link between microscopic structural features (well-known as descriptors) of materials and their macroscopic properties (this can be any property such as density, viscosity, toxicity, *etc.*). The following general form can be considered for the property model:

$$\text{Property} = f(\text{parameters/descriptors}) \quad (1)$$

In the case of empirical and semi-empirical models, parameters/descriptors that are used to obtain the model are very important and their selection is a crucial task. Depending on the approach different types of descriptors can be considered. These descriptors are obtained experimentally, theoretically, quantum-mechanically (chemically) (QM) or molecular mechanically (MM) and a combination of all types of descriptors. Having access to high accuracy experimental database is necessary. Some examples of these data are experimental values reported in the literature, or famous databases like Design Institute for Physical Properties (DIPPR),[64] NIST,[65] and DETHERM.[66] Poling *et al.*[67] notes there is a relation between molecular structure and the bonds between atoms and their macroscopic properties. This concept proposes that a macroscopic property could be estimated using group contribution (GC) models. GC models include a wide range of models such as activity coefficient GC models like UNIFAC to EOS GC models like SAFT.[68]

Quantitative-structure property/activity relationship (QSPR/QSAR) is a modelling method to predict different physical and thermodynamic properties using the knowledge about the chemical structure of the molecules.[69] These physio-chemical structure and properties are known as descriptors and provide the basis for mathematically linking and explaining a molecules/materials activity or property. A large family of models have been developed to predict the properties for solvent-based CO_2 capture systems based QSPR approach. Different modelling (regression) approaches are applicable in QSPR/QSAR studies which are different from linear techniques like multivariate linear regression (MLR), partial least-squares regression (PLSR), and

principal component regression (PCR) to the nonlinear techniques such as ANN, GP, SVMs, and ANFIS. In QSPR studies especially when dealing with MLR method, different types of algorithms from classic algorithms such as stepwise forward selection or evolutionary or metaheuristic algorithms such as genetic algorithm (GA), particle swarm optimization (PSO), simulated annealing, and ant colony and so on, have been used in descriptor selection step to reduce the number of descriptors and keep the most influential ones in the prediction of property under study.

3.1.3.2 Review of the ML-based property modelling studies. Many of the descriptors that have been used in QSPR models related to CO_2 capture have physical meaning. Temperature, pressure, partial pressure of CO_2, the concentration of solution are some examples of descriptors that are used by different researchers. Golzar *et al.*[70] developed ANN QSPR model to predict the solubility of CO_2 and N_2 in common polymers. The authors used genetic function approximation (GFA) to find the best descriptors between 1600 molecular descriptors. They found out that molecular weights of gas and monomer, spectral moment 07 from edge adj. matrix weighted by edge degrees, mean atomic Sanderson electronegativity (scaled on Carbon atom), mean atomic polarizability (scaled on Carbon atom) can predict the solubility of CO_2 and N_2 in common polymers. Venkatraman and Alsberg[71] extracted over 10 000 IL-CO_2 solubility data for 185 ILs measured at different operating temperatures and pressures from the literature. The authors used a single decision tree, PLSR, and the non-linear ensemble random forest models. They also considered the COSMO-RS model and predicted the results of regression models with this quantum mechanical based thermodynamic model. They reported that temperature and pressure and parameters relevant to intermolecular interactions were selected as descriptors of the models. In this regard, a number of HOMO, LUMO energy-based descriptors such as the HLFRACTION (ratio of the HOMO/LUMO energies), softness (inverse of the HOMO–LUMO gap) is indicative of the cation–anion electrostatic (nucleophilic–electrophilic) interactions that are key to the CO_2 solvation abilities to be selected. Other descriptors focus on important geometrical parameters such as the ovality or its inverse, the globularity factor that reflects the ability of the molecule to adapt its shape with respect to the approaching reactant. Kuenemann and Fourches[72] collected and compiled experimental absorption properties for more than 40 unique amines, and developed several QSPR models demonstrating the influence of structural modifications for amines' absorption properties. The authors used different MLs techniques namely ensemble tree, partial least squares regression, random forest, and ANN. They reported that the Random Forest and ANN models gave the best results. The authors also mentioned that they considered two types of descriptors in their study namely RDKit descriptors and Functional Connectivity Fingerprints (FCFP). A total of 117 RDKit and 1024 FCFP6 descriptors were computed. After pre-treatment of data, their dataset of amines reduces to 67 RDKit descriptors and 140 FCFP6 fingerprints descriptors. Zhang *et al.*[73] used ANN to predict CO_2 solubility in the solutions of potassium lysinate (PL) and its blended solutions with MEA, with a total of 433 data groups extracted from the literature. They use two different methods namely BPNN and general regression neural network (GRNN). The authors also predicted the aqueous solution density and viscosity using the same method. Afkhamipour *et al.*[74] selected concentration, temperature, molecular weight and CO_2 loading of the amine as the inputs (descriptors) to the ANN model to predict the heat capacity (C_P). Here, 3947 experimental data points representing heat capacity for 47 systems of amine-based solvents with a broad range of concentration and temperature were collected from published papers. The AARD% between model results and experimental data of C_P for amine-based solvents was 4.3%. The obtained results from the ANN and thermodynamic models showed that the models could accurately predict the C_P of conventional amines with an AARD% of 0.59%, and 0.57%, respectively. Cao *et al.*[75] modelled the toxicity of ILs towards a leukaemia rat cell line (ICP-81) using QSPR method. The authors considered the structures of 57 cations and 21 anions that were optimised using quantum chemistry. The ML methods used in this study were extreme learning machine (ELM), MLR and SVM. The results show that the ELM method had the best statistical parameters. In the aspect of used descriptors in their model, Sσ-C-0.016 stands for the charge distribution area of the cation. SEP-A-69.25 and SEP-A-128.75 belong to the electrostatic potential surface area of anions. The other selected descriptors in their model are related to the electrostatic potential surface area of cations. The authors emphasised that the parameters for the electrostatic potential surface area are important and effective descriptors for predicting the toxicity of ILs. Borhani *et al.*[76] used GA-MLR method to develop a model to predict the partial pressure of CO_2, the heat of absorption, and *K*-values for CO_2 absorption in 30, 45, and 60 wt% MEA aqueous solutions. The GA was used for the selection of the best parameters (feature selection) and functional form, by optimising with respect to the RQK fitness function. They used combination of CO_2 loading and temperature as descriptors to predict the partial pressure of CO_2. Mazari *et al.*[77] predicted CO_2 solubility, density, viscosity and molar heat capacity of an IL ([Bmim][PF_6]) using three GPR family and SVM methods. The range of temperature, pressure and water content of the data used in the models are presented in Table 3.

The results showed that the least accurate model was SVM with an AARD% of 15.13. The squared exponential GPR model was the most accurate coefficient of determination of 0.992 and AARD% of 0.14 for testing data. Wu *et al.*[78] collected a total of 160 experimental data points for Henry's law constant of CO_2 in 32 imidazole ILs. Multi-Layer Perceptron (MLP), RF and MLR were used to develop the models to predict Henry's law constant. The results of the modelling showed good statistical parameters for all three models for the test set. The correlation coefficient mean absolute error (MAE), and RMSE for the MLP model were 0.98, 0.4818 and 0.65 respectively. The authors considered temperature, CO_2 partial pressure and water wt% as input of the model (descriptors) which all of them have physical meaning here.

Table 3 The range of experimental data extracted from the literature to predict different thermodynamic properties of ionic liquids[77]

Dataset	T/K	P/MPa	Water/wt%
CO_2 solubility (mass fraction of CO_2)	293–395	0.015–9.685	0–1.6
Density (g cm^{-3})	278–391	0.1–173	0–2.68
Viscosity (mPa s)	273–388	0.1–175	0
Heat capacity (J K^{-1} mol^{-1})	283–353	0.1–100	0

3.1.4 Solvents selection and design

3.1.4.1 Background and challenges of solvent selection and design methods. Two important methods can be used to screen the solvents to absorb the CO_2-application of chemometric models (QSPR, GC,...) and computer-aided molecular design (CAMD).[79] Since ML is utilised in both these types of methods, it can therefore be said that the models developed using ML and described in Section 3.1.3 can be used to screen and select the best solvents.

3.1.4.2 Review of the ML-based models used to solvent selection and design. ML has been used to perform solvent screening for different applications.[80,81] Some studies related to solvent screening for CO_2 absorption have been done using the COSMO-RS thermodynamic model.[82,83] However, it should be noted that the number of studies related to the application of ML in the solvent selection and design for CO_2 absorption is considerably less than the application of ML in other types of studies related to CO_2 absorption, which are reviewed in previous sections. As ML is used to select and screen solvents for different applications,[81] it is promising to use it for CO_2 absorption solvents as well. Venkatraman *et al.*[84] have employed a multi-property, high-throughput pipeline to facilitate task-specific IL discovery. In Fig. 3, one of the main steps is the application of ML. ML models (RF, cubist and gradient boosted regression (GBR) were developed using experimental data for 10 different IL properties of interest. The models were applied to a large library of eight million cation–anion pairs that span diverse chemical scaffolds.

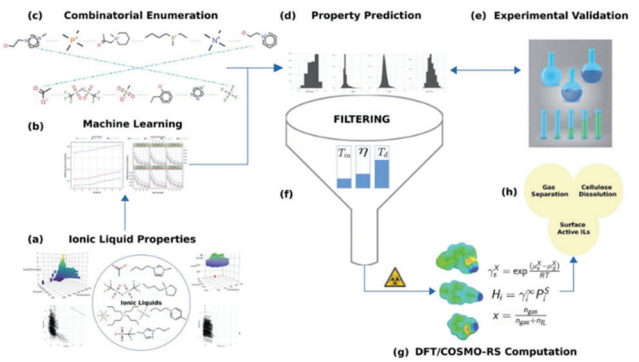


Fig. 3 The concept of the approach presented by Venkatraman *et al.*:[84] (a) data collection, (b) ML calibration, (c) combinatorial library design and enumeration, (d) prediction of properties by ML, (e) experimental validation of selected candidates, (f) property-based filtering, (g) theoretical evaluation, (h) potential applications.

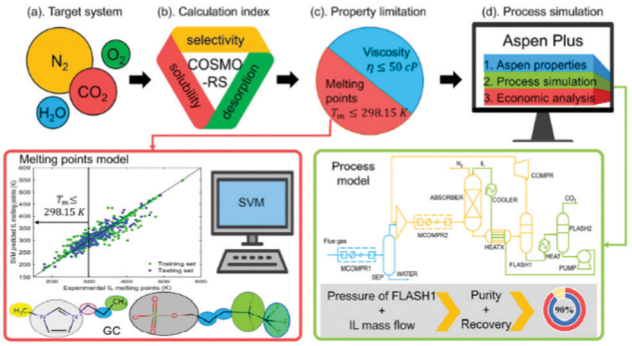


Fig. 4 Strategy considered to select and evaluate the best candidates of ILs.[85]

Wang *et al.*[85] presented a strategy to select the best ionic liquids and apply them in the process simulator to absorb CO_2. Their strategy contains four main steps. The first part is related to the target system, in the second part absorption, selectivity and desorption for each IL are calculated using the COSMO-RS model. In the next step, a prediction model is applied to predict viscosity and another one for predicting melting point to find the optimal ILs which these models are developed using the SVM method. In the final step, the applicability and effectivity of optimal ILs reported in the literature are evaluated by Aspen Plus® (Fig. 4).

3.1.5 Perspectives and prospects. In comparison to the first principle models developed for different studies on CO_2 absorption, the ML models are more accurate as they provide a complex and non-linear relationship between the inputs and predict the targets. As noted in this section, many ML-based models are developed for different applications of CO_2 absorption. However, the models that were developed for the prediction of physical and thermodynamic properties were not applied in any process modelling study. An important future goal is to integrate MATLAB or Python or other similar ML programmes to Aspen Plus® or gPROMS® or similar simulators to use these ML-based models in first principle process modelling studies. As ML-based models are more accurate than the traditional models, they can result in better predictions and results in thermodynamic and process modelling studies. Hence, the connection of these models to process simulators should be considered in future studies.

3.2 Machine learning in CO_2 adsorption

Adsorbents are micro-porous structures with a characteristically large surface area and the ability to capture large amounts of gases on their surface.[86] They generally have a selective affinity for specific gases in a mixture of gases, making them ideal for gas separation applications such as CO_2 capture.[86,87] One of the primary considerations when designing an adsorbent-based CO_2 capture process is the choice of the adsorbent media.[16] This field has gone through a renaissance in recent years with the advent of the use of organometallic chemistry.[24,88,89] There are several new classes of adsorbents such as Metal–Organic Frameworks (MOFs),[88–90] Covalent

Organic Frameworks (COFs),[91] Zeolitic Imidazolate Frameworks (ZIFs),[92] Porous Organic Cages (POCs)[93] along with the classical zeolites[86,94] and activated carbons. Many of these porous structures are chemically and physically tuneable and can be reverse engineered to provide the process designer with tailor-made options.[95–97] This means an effectively infinite number of possible structures can be theorised.[24,98] Exploring the entire adsorbent material design space is computationally restrictive, and traditional adsorbent characterisation techniques are time-consuming, adding to the complexity.[98,99] Large databases with over one million of such real and *in silico* hypothetical porous structures are available to process designers that are already partially characterised for the application of CO_2 capture.[24,100–106]

3.2.1 Adsorbent synthesis and characterisation. The ability to build adsorbent structures by using a different set of building blocks has been well documented in the literature.[107,108] This has provided a realistic opportunity to tailor-make an adsorbent for CO_2 capture with targeted features such as high CO_2 affinity over other gases in the flue gas mixture.[16] However, with an almost infinite set of possible structures, correctly identifying the best adsorbent is extremely challenging. To make matters more complicated, the required adsorbent properties for an effective CO_2 separation process are not fully understood.[24,25,109–112] The discovery and synthesis of new adsorbents using traditional experimental techniques alone are expensive and time-consuming.[113] Computational methods have been used to create frameworks to develop, characterise, and tune the properties of the porous structures.[97,100,114] ML *via* supervised and unsupervised algorithms can help explore the complex and highly multivariate material design space.[24,98] Researchers have already applied many ML and other statistical techniques to explore adsorbent synthesis pathways.[99] Other aspects for adsorbent selection for applications such as CO_2 capture are the synthesizability, stability to moisture, and overall life cycle costs, among other things, which can be aided by the application of ML.

Adsorbent discovery and screening for CO_2 capture using supervised ML models have been extensively reported in the literature.[99] There have been many instances in the literature where the adsorbent properties are also tuned for specific applications. Collins *et al.*[115] showed that a genetic algorithm could efficiently optimise for desired physical or functional property in MOFs by evolving the functional groups within the pores. The authors optimised the CO_2 uptake capacity of 141 experimentally characterised MOFs under post-combustion CO_2 capture conditions and were able to increase the CO_2 adsorption on MOF MIL-47 by 400%. ML models have also been used to identify novel adsorbent properties such as hydrophobic adsorbaphore. This could be a very interesting phenomenon to exploit since the presence of moisture always hindered adsorptive CO_2 capture. Boyd *et al.*[116] screened an adsorbent library of $\approx$300 000 structures to identify adsorbents with this adsorbaphore property and demonstrated a synthesis pathway for two such adsorbents. These demonstrations of ML in the discovery, synthesis and exploration of the adsorbent design space show the possible pathways for identifying and implementing an effective adsorbent-based CO_2 capture process.[116]

ML techniques have also been applied to speed up the characterization of the adsorbents. The Grand Canonical Monte Carlo (GCMC) is generally used to predict the adsorption, and Molecular Dynamics simulations (MD) are used to describe diffusion and other transport properties.[117,118] These techniques have been used to generate adsorbent property data for large databases of adsorbents at enormous computational costs.[105,119] To tackle this problem, researchers have applied supervised ML techniques to build predictive data-driven models. Extensive work has been carried out by computational materials chemists to identify the underlying QSPR using ML.[120] There are four general classes of descriptors that are generally used to describe the adsorption equilibria, geometric, topological, chemical and energy-based.[121] Dureckova *et al.*[122] developed ML models to predict CO_2 working capacity and CO_2/H_2 selectivity using a diverse set of MOF structures using gradient boosted trees regression method. The authors also showed that both geometric descriptors, such as surface area, and chemical descriptors, constructed using atomic property weighted radial distribution functions, can be used to predict with reasonable accuracy the working capacity and mixture gas selectivity.[122] Burner *et al.*[123] presented a similar framework to predict the working capacity and CO_2/N_2 selectivity using a deep neural network (DNN). The best predictions were obtained with the AP-RDF, chemical motif, and geometric descriptors, all as inputs, with an $R_{adj}^2 > 0.95$. Pardakhti *et al.*[124] reported that a framework for the prediction of methane uptakes using ML algorithms. They evaluated multiple ML algorithms, such as SVR and RF, and reported a high prediction accuracy compared to the GCMC predictions.[124] Bucior *et al.*[125] presented a data-driven surrogate trained ML model to predict H_2 loading on MOFs using a new type of descriptors as model inputs. The descriptors were derived using the binned histograms of the energies of adsorbent–adsorbate interaction and used as inputs to the predictive model. The sparse regression model trained with this and geometric descriptors to predict gas uptake in multiple MOF databases to a high degree of accuracy.[125] These studies show us that both the adsorbent structure and the chemical interactions are needed to be taken into account for accuracy in predictions. ML frameworks have been successfully shown to speed up single adsorbent–adsorbate interactions. Still, their real application is in the prediction of multiple gases and mixture gas adsorption on adsorbents. Techniques such as transfer learning, dimension reduction, feature identification can improve the model predictions for such cases.[126] Anderson *et al.*[127] presented a new framework to predict the adsorption of multiple adsorbate gases for a given range of conditions using a MLP. The model was trained using the variables that describe the force-field parameters of ''alchemical'' species and the MOFs as simple descriptors such as geometric and chemical moieties. The resulting models could then predict the adsorption of six different gases in a diverse set of adsorbents.[127]

While understanding the separation potential of an adsorbent is critical, quantification of the mechanical stability and synthesizability of the *in silico* predicted adsorbent structures is an important aspect for the final deployment of the technology.

Evans *et al.*[128] showed that ML models predicted bulk and shear moduli of zeolites using only geometric features and that the accuracy of these predictions is better than the traditional force field approaches. Moghadam *et al.*[113] demonstrated that ML techniques and multi-level simulations predict MOF properties. The ML models developed in this work can predict the mechanical properties of MOFs in a matter of seconds. They were also shown to predict the mechanical stability for the *in silico* predicted structures.[113]

The recent explosion of ML-related applications means that a large amount of new information, through publicly shared models and data, open up the possibility of transfer learning. Here, models taught to learn patterns for a specific application or purpose can help retrain new models for different applications. This has been demonstrated for applications such as the characterisation of adsorbent isotherms, where ML models used to predict equilibrium measurements of one gas can help the prediction of other gases on the same adsorbent. Thus, saving precious computational time.

3.2.2 Process modelling and optimisation. Cyclic adsorption processes are typically operated in fixed beds that undergo several steps to achieve the desired separations. Depending on the bed regeneration strategies, several processes operational modes such as pressure swing adsorption (PSA), vacuum swing adsorption (VSA), temperature swing adsorption (TSA), temperature-vacuum swing adsorption (TVSA), concentration swing adsorption (CSA), electric swing adsorption (ESA), microwave swing adsorption (MSA), *etc.* can be realised. Such systems are inherently characterised by a system of coupled nonlinear PDEs obtained from the underlying mass, momentum and energy balances. In the context of modelling and simulating cyclic adsorption processes, the system of nonlinear PDEs is repeatedly solved in time and space for each step in a cycle sequence. Owing to its transient and cyclic nature, adsorption processes must be simulated until the system reaches a cyclic-steady state (CSS). The key performance indicators are then calculated based on the transient profiles of state variables (composition, pressure and temperature). Often, solving the system of PDEs cyclically several times until CSS is computationally demanding. Further, the modular nature of cyclic adsorption processes allows for flexibility in controlling several operating conditions and design parameters. Hence, in the context of process optimisation, several decision (or design) variables can arise. Therefore, the high-dimensionality and effort to determine process performance at CSS make optimisation of cyclic adsorption processes complex and challenging.

To tackle problems mentioned above, ML techniques have been applied to design and optimise cyclic adsorption processes for CO_2 capture applications. The studies employing ML to model and optimise cyclic adsorption processes can be classified into three categories. The first category corresponds to studies that used ML for supervised learning (regression) to know the structural mapping between the decision variables and process outputs in the process optimisation in order to avoid the computational burdens of running high-fidelity simulations for functional evaluations. To this end, an initial design of experiments (DOE) is performed on the decision variables that typically cover the entire design space. The high-fidelity models are then used to calculate the desired process outputs (typically key performance indicators used in the optimisation) based on the sample set of decision variables from the DOE. Finally, surrogate models using ML algorithms are constructed based on those samples and subsequently used in the optimisation. Single or multiple surrogate models can be constructed for process outputs. For example, Pai *et al.*[129] tested the ability of a variety of surrogate models constructed based on different supervised ML algorithms to predict the performance indicators of a 4-step VSA process for post-combustion CO_2 capture. Algorithms such as decision trees, RFs, SVMs, GPR and ANNs were trained for each performance indicator using a sample set of operating conditions generated *via* Latin hypercube sampling. Among these, GPR was shown to perform well using an adjusted coefficient of determination (greater than 0.98) as the metric. Upon employing these surrogate models in the process optimisation, they showed that the relative error of the optimal performance indicators from the surrogate and high-fidelity simulations was within 3%. Subraveti *et al.*[130] developed a neural network-based optimisation approach to determine the Pareto solutions of multi-objective maximisation of CO_2 purity and CO_2 recovery for a complex 8-step PSA process designed for pre-combustion CO_2 capture. Herein, the multi-objective NSGA-II (Non-Dominated Sorting Genetic Algorithm version II) algorithm's initial generations were carried out using high-fidelity simulations for evaluating objectives. This also served as the training data generation step for the neural network models, which learned the underlying input–output mapping structures between decision variables and objectives, CO_2 purity – CO_2 recovery. Such training data that was already biased towards the optimal region of the decision variable space helps improve the prediction accuracy of the neural network models in the desired optimal region. A three-layer feed-forward neural network with one input layer, one hidden layer with ten neurons and one output layer were used for each objective to demonstrate this approach, with results indicating that the relative error in both the objectives was found to be around 1%. The PSA optimisation using neural networks was ten times faster as compared to using high-fidelity simulations for functional evaluations. Instead of constructing a surrogate model for each performance indicator, Xiao *et al.*[131] used a multi-output feed-forward neural network architecture to predict purity, recovery and productivity in the PSA optimisations. Vo *et al.*[132] formulated an integrated process model based on the combination of different feed-forward neural networks, which represent the input–output mapping structure of cryogenic, membrane and PSA units for hydrogen recovery and CO_2 capture from the tail gas of SMR-based hydrogen plants. The neural network models for each unit were shown to have less than 2% error and were subsequently used to minimise the production cost of the integrated process. The neural network models were also shown to have low computational costs.

Often, uncertainty arises in ML-based optimisations during the ML model selection and/or training the model parameters.

Uncertainties in model predictions even lead to potentially different optimal solutions. To address the issue of uncertainties in ML-based optimisations, Hüllen *et al.*[133] proposed three different strategies, *i.e.*, robust optimisation, stochastic programming and discrepancy modelling, integrated with ML models for handling uncertainty. These approaches have been applied to a case of temperature swing adsorption process for DAC where the productivity of the process was maximised subject to purity, recovery and energy constraints. Sparse Grid polynomials and ANNs were used as data-based models to approximate decision variable-processes output mapping. The authors stress the importance of incorporating uncertainty into ML-based optimisations.

The second category of studies involves developing supervised ML models to predict the axial or temporal profiles of the cyclic adsorption process. Pai *et al.*[129] also developed neural network models to predict the bed profiles of the intensive variables of a 4-step VSA process at CSS. Using these neural networks, they demonstrated a rapid convergence to CSS. Further, the neural network predictions were also matched with the experiments. Leperi *et al.*[134] used neural networks to construct basic steps in typical PSA processes for post-combustion CO_2 capture. For each step, twelve neural network models were constructed. To elaborate, each neural network model for predicting five state variables (absolute pressure, CO_2 gas phase mole fraction, CO_2 molar loading, N_2 molar loading and column temperature) were measured at ten measured locations along the column. Further, one neural network at each end of the column predicts the total gas flowing in and out of the column. This approach allowed them to synthesise different PSA cycles for post-combustion CO_2 capture and calculate their performances based on the neural network models underpinning each step. Oliveira *et al.*[135] proposed a real-time soft sensor for a PSA unit based on deep learning networks. Three different types of ANNs, namely, feed-forward, recurrent and long short-term memory (LSTM) models based on multi-input and a single output, were developed to predict the PSA model dynamics. It was shown that LSTM-based DNNs outperformed feed-forward and recurrent neural networks in terms of predicting the dynamics of PSA. The authors also suggested that the LSTM-based DNNs can be reliable for optimisation, control and on-line measurements of PSA units.

In the third category, supervised ML algorithms such as PLSR were used for reducing the dimensionality of the cyclic adsorption process optimisation. For example, Subraveti *et al.*[130] employed PLSR to identify each decision variable's relative importance in the optimisation, which impacts the process objectives. The most relevant decision variables were identified using the PLS weights, and other variables are discarded. For the case study considered, the original eight decision variables were reduced to three using this approach. This improved the optimisation speeds by almost 50% without compromising the accuracy of the Pareto solutions.

3.2.3 Integrated material-process screening studies. The choice of the porous adsorbent media is dependent on the product requirements and constraints. Traditional adsorbent selection metrics such as selectivity, and working capacity, fall short of this and thus do not provide the complete representation of separation efficiency/performance.[136] Additionally, many such simplified metrics do not fully consider the process requirement or the complex multiscale phenomenon during scale-up. Although relevant and valuable work has been carried out in relation to the underlying QSPR in most of cases, there needs to be a consensus over the integration of the real-world process that will be used to separate and capture the CO_2.[137] Often, simplified descriptors such as CO_2 working capacity or selectivity are used as optimisation targets.

ML-based techniques such as DNNs are well-suited for applications that require large amounts of repetitive computation. ANN-based surrogate models have been applied as cheap computational emulators of complex process models to aid in the fast screening of material. Khurana and Farooq[111] developed regression models to directly predict minimum energy and maximum productivity for CO_2 capture from a flue gas stream containing 15% CO_2 using a VSA process. Khurana and Farooq[111] also screened around 80 adsorbents using the ML model and validated the optimised results with a detailed mathematical model. Burns *et al.*[25] and Leperi *et al.*[110] also screened the CoRE MOF database to identify high-performance adsorbents for post-combustion CO_2 capture using a detailed model. Burns *et al.*[25] developed a decision tree-based ML model, and Leperi *et al.*[110] developed a generalised separation metric using the data from a detailed model to screen new adsorbents in the same process with a high degree of accuracy. These papers also showed the clear computational advantage of the application of ML-based surrogate models for screening due to their inherent speed and accuracy. Pai *et al.*[26] developed a generalised framework called machine-assisted adsorption process learner and emulator (MAPLE) for modelling and screening any Langmuir (Type I) adsorption isotherm by including the isotherm parameters as model inputs along with the process parameters. The authors demonstrated that the framework accurately modelled process performance and were able to validate the ML-based optimisation framework from the external literature. The study showed the computation required to train the generalised ML model was similar to the computation required to screen $\leq$ten adsorbents using the traditional modelling and optimisation approach. It should be noted that these ML models are robust only in the training data range. One must be careful not to overtrain and to thoroughly validate the performance with independently generated testing data.

3.2.4 Process inversion and performance limits. In recent years, multiscale models have shown that it is necessary to carry out the integrated process and material screening.[25,110–112] However, the full consideration of all the multiscale phenomenon makes the computational evaluation restrictive. For this reason, most scale-up studies in the literature evaluate only a small subsection of the available adsorbents. This makes effective and accurate screening of adsorbents a non-trivial problem. Alternatively, reverse engineering the hypothetical best performing adsorbent for a fixed process cycle, where the operation of the process cycle is optimised, is a route to identify the

best possible choice, with the final goal being the synergistic design of both the adsorbent media and separation process cycle. In each of these cases, vast amounts of simulation experiments need to be carried out.

Khurana and Farooq[111] developed an inverse design framework to predict the hypothetical best isotherm for post-combustion CO_2 capture in a VSA-based process. In this work, the authors considered five input parameters to describe the adsorption equilibria and trained a neural network model. The resulting optimisation of the idealised isotherms provided insight into the effect of the isotherm on the process performance. Pai *et al.*[137] used a ML surrogate, MAPLE, for a wide range of operational conditions and used the inverse adsorbent design approach to study the limits of PVSA-based CO_2 capture for a wide range of CO_2 feed compositions. Yao *et al.*[138] proposed an automated adsorbent discovery framework using an auto-encoder to generate MOF structures with desired functions. The results showed that the model accurately captured structural features and was able to reconstruct MOF structures. The framework showed the automated design of MOFs for CO_2 capture from natural gas and flue gas streams.[138] These studies highlight the advantage of ML in synergistic processes and adsorbent. Due to their computational speed and accuracy, such ML models allow designers to explore previously computationally restrictive engineering problems.

3.2.5 Perspectives and prospects

Material design and discovery. The material databases include more than 500 000 structures (both experimental and hypothetical) that can be evaluated for CO_2 capture. Such large databases can be screened for best performers using ML. Unsupervised/semi-supervised learning methods can be applied to classify the materials in databases into different clusters and know the underlying patterns/distributions within the databases. In addition, supervised learning techniques can be used to identify the mapping between the structures and material properties without the associated computational burdens of solving physical models.

Process modelling and optimization. The major barrier for exploring different adsorption process cycles for CO_2 capture has been the significant computational demands in process modelling and optimisation. Existing studies in the literature showed that supervised learning algorithms could be efficiently incorporated into the optimisation routines. With the advances in ML, more efforts must be directed towards the dynamic modelling of adsorption processes. For instance, Leperi *et al.*[134] used ANNs to model the dynamics of some basic constituent steps in PSA processes. Such approaches are useful, especially when designing and evaluating different adsorption processes for CO_2 capture. Increasing the generalisation capability of such ML models is also important for accurate predictions. These models can also gain more insights in understanding the interplay among different intensive variables such as gas composition, pressure, temperature, and solid compositions affecting the process. The high dimensionality of the adsorption process optimisations can be tackled using ML. Semi-supervised/unsupervised algorithms can be utilised to know the effect/causal relationships between the decision variables and the performance indicators. This will help understand the underlying relationships between process inputs-outputs and identify significant decision variables for the optimisation. While most ML studies are focused on the processes designed for the pilot-scale, some of these ML approaches can also be extended to industrial applications. For example, these models can be effectively used in the process monitoring and control to overcome inherent process control challenges, especially since several sequences of steps occur in cyclic adsorption processes. Reinforcement learning (RL) can also be applied to monitor and control the cyclic adsorption processes. RL algorithms can be trained to learn adaptability when the process is subjected to external disturbances.

Integrated material-process screening. For CO_2 capture, integrated material-process studies have recently become common. Given that a large number of materials have to be screened using the process for reliable material evaluations, conducting a multiscale computational campaign for integrated material-process performance evaluation is computationally very expensive. However, ML has transformed this potentially computationally impossible exercise into a possibility. For example, Pai *et al.*[26] developed a material agnostic ML framework where both material and process decision variables are considered for screening and evaluating the performance of different materials. Such approaches will enable a deeper understanding of the underlying patterns in the material feature space. Algorithms like manifold learning can be utilised to identify such patterns in the material feature space, which will help in accelerating the material discovery for CO_2 capture.

3.3 Machine learning in oxy-fuel and chemical-looping combustion for CO_2 capture

3.3.1 Machine learning in oxy-fuel and chemical-looping combustion. Oxyfuel combustion burns fuels in a mixture of pure O_2 and recirculated CO_2 instead of air, and then the CO_2 can be easily separated from the flue gases. To reduce the energy penalty and costs from the air separation unit in the oxy-fuel combustion process, the next generation of carbon capture technology, chemical-looping combustion (CLC), that can transfer the oxygen from the air reactor to the fuel reactor by means of oxygen carriers, has been proposed. The current technology readiness level (TRL) for oxy-fuel combustion and CLC is estimated at 7–8 and 6, respectively. The applications of ML in these technologies are mainly focused on predicting the thermodynamic characteristics of oxy-fuel combustion, monitoring the oxy-fuel combustion process, estimating the reactivity of oxygen carriers and process control of CLC.

To reduce the complexity and improve the accuracy of numerical models to predict the coal/char combustion rates, Zhu *et al.*[139] investigated the application of an ANN approach for estimating the coal/char combustion rates with their characteristics as inputs of the neural networks. The results indicated that ANNs can provide a new approach to the development of models for predictions of reactivity/combustion rate of coal

combustion with reasonably good accuracy and robustness.[139] Later on, several researchers employed ANN to predict the values from thermogravimetric analysis (TGA) of oxy-fuel combustion of different fuels. Chen *et al.*[140] applied ANN models to predict the thermogravimetric curves of co-combustion of sewage sludge and coffee grounds under O_2/CO_2 atmospheres, with O_2/CO_2 mixing ratios, heating rates, and temperature as the inputs. After training using the experimental data from the TGA, the optimal ANN model provided a good agreement between the experimental and predicted values. Xie *et al.*[141] compared the performance of RBF and BPNNs on the prediction of TG curves of oxy-*co*-combustion of textile dyeing sludge and pomelo peel, with the mixing ratio, heating rates, combustion atmosphere and temperature as the inputs and mass loss percent as the output. The results indicated that BPNNs gave a better prediction than that of RBF neural networks.[141] Govindan *et al.*[142] used trained ANNs, using TGA to predict the sample mass loss percentage of oxy-fuel combustion of calcined pet coke, with the predictions obtained from the model showing a high degree of accuracy, with a coefficient of determination (R^2) of 0.99. Qiao and Zeng[143] also applied the ANN framework to predict the gas products of heavy oil gasification under oxy-fuel conditions but the authors have not clarified how they trained and validated their ANN models. Debiagi *et al.*[144] developed a reduced-order model based on ML, which can accurately predict different phases of coal particle combustion at a reduced computation cost. They used a High Dimensional Model Representation (HDMR) method to develop the supervised ML models (see Fig. 5). Unlike the case with the previous work, the training and test datasets were generated from an accurate, detailed solid fuel kinetic model that considered a wide range of operation conditions obtained from a novel gas-assisted coal combustor.[144]

Krzywanski *et al.*[145] developed a generalised ANN model to predict the SO_2 emissions from large- and small-scale circulating fluidised bed (CFB) boilers under air-firing, oxygen-enriched and oxy-fired combustion conditions with the dimension and operating parameters of the CFB boilers as the inputs. The authors[145] also conducted a sensitivity analysis to investigate the effects of changing operating parameters on the SO_2 emissions using the trained ANN models. The results indicated that the ANN model can serve as a fast tool to provide the accurate prediction of SO_2 emissions for coal combustion in the CFB boilers under the different combustion environments with less complexity and costs.[145]

Besides predicting the useful parameters of oxy-fuel combustion, ML can also be applied to monitor air/oxy-fuel combustion processes for combustion control and optimisation under variable conditions. Bai *et al.*[146] proposed a novel method by combining flame imaging, principal component analysis and random weight network (PCA–RWN) techniques for multi-mode process monitoring for air and oxy-fuel combustion of coal (see Fig. 6). Flame image database collected from a 250 kW air/oxy-fuel combustion Test Facility were used to validate the PCA–RWN models and the performance was evaluated by the Hotelling's T^2 and squared prediction error (SPE). Compared to the performance of the proposed PCA–RWN model with other ML classifiers (Kernel Support Vector Machine, Neural Network, and *k*-Nearest Neighbour classifier) for pattern recognition, the proposed PCA–RWN model gives the best prediction of the average recognition success rate and the least training time.[146] The authors[147] also followed a similar methodology to apply the PCA with kernel support vector machine (KSVM) model for the multimode monitoring of combustion stability under different oxy-gas fired conditions. Liu *et al.*[148] used a supervised multilayer deep belief network (DBN) to evaluate the nonlinear relationship between the flame images and the outlet oxygen content, and the results indicated that the proposed method was a reliable and efficient way for predicting the real-time oxygen content. Later on, Han *et al.*[149] applied flame imaging and stacked sparse autoencoder based DNN to monitor the combustion stability. The results showed that the proposed model could quantitatively and qualitatively evaluate the combustion stability with good generalisation and robustness.[149]

Yan *et al.*[17] used the experimental data of nineteen manganese ores to train the ANN models to predict the reactivity of manganese ores as oxygen carriers in CLC. The results indicated the optimal ANN models can provide very good performance predictions for both training and new dataset and the authors proposed a general workflow in applying ML model to predict the performance and aid the design of the oxygen carriers as shown in Fig. 7.

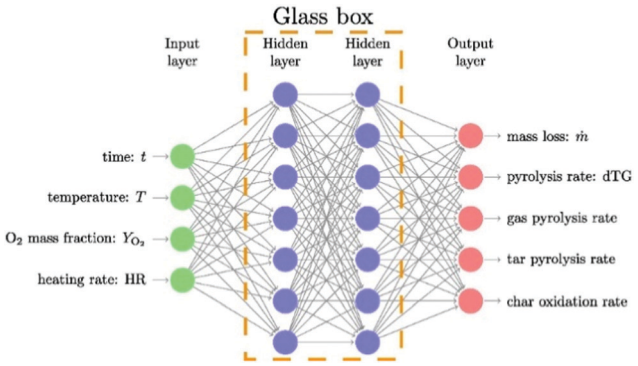


Fig. 5 Diagram of a generic multilayer perceptron of the HDMR method.[144]

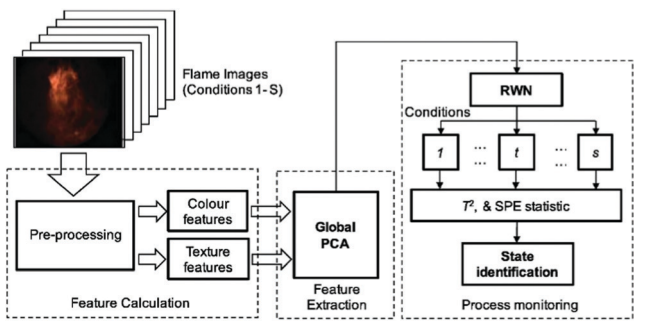


Fig. 6 Diagram of PCA–RWN model for multi-mode combustion process monitoring.[149]

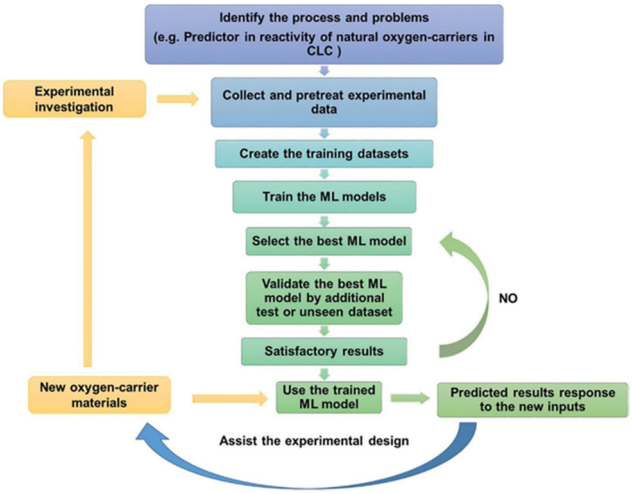


Fig. 7 Workflow of developing a machine-learning model for oxygen carriers in chemical-looping processes.[17]

Singstock *et al.*[150] proposed a statistical ML descriptor-based method to predict the reaction free energies and classify the thermodynamically viable active materials for chemical-looping processes, and the authors applied it to evaluate materials for a novel chemical looping process for pure SO_2 production. This approach is envisioned to link the process design with high-throughput material discovery to promote the development of a wide range of chemical-looping technologies.[150] Wilson and Sahinidis[151] proposed a mixed-integer nonlinear programming (MNLP) formulation to estimate and identify kinetic rate parameters from a postulated superset of reactions, and they validated that this approach can automatically generate accurate kinetic models from dynamic CLC process.

The assurance of smooth and long-term operational stability of the CLC system is one of the key requirements for CLC technology to be deployed on a commercial scale. Pan *et al.* applied the LSTM based recurrent neural network (RNN) for early detecting of fault caused by fines accumulation, which is represented as bubbles in the packed bed standpipe of a chemical looping systems. The results revealed that the model trained by the cold-flow model of sub-pilot scale chemical looping system can provide a recall value of at least 86.7% with the application of ensemble decision strategy, and the authors pointed out the proposed model can easily be extended and generalised with further training using the data obtained from multiple operation conditions.[152]

3.3.2 Machine learning in calcium looping. A similar process to chemical looping, is calcium looping, which is a CO_2 capture process, that uses calcium oxide-based sorbents to separate and remove CO_2 from flue gases. The process is based on the reversible reaction of lime with CO_2 and is considered as an emerging CO_2 capture technology. This process has been well researched with findings focusing on optimal CaO based sorbents to achieve the best capture efficiency, however the application of ML to this field is relatively new, with very few studies on this aspect.

Chen *et al.*[153] proposed the use of BPNN to predict the performance of Ca-based sorbents in the calcination/carbonation cycles, based on TGA experimental data. This study observed the factors that affected the sorbent performance, namely sample particle diameter, calcination temperature, calcination duration, calcination atmosphere and carbonation duration. The feed-forward multilayer ANN, which had the architecture of 5-34-1, had the five aforementioned factors as inputs, and the carbonation conversion degree as the output parameter, calculated with the assumption that the decomposition of calcium carbonation was the only reason for sample weight change. Here, 75% of the data was used for training while the remaining 25% was accounted as the test data. The model proposed showed a strong correlation with TGA results and proved the validity for the approximation of Ca-based sorbent in the carbonation process even when conducted at extreme reaction condition.

A recent application of ML to the calcium looping process was developed by Nkulikiyinka *et al.*[154] Here, the authors developed an ANN and random forest (RF) model to act as soft sensor models, for the prediction of gas concentrations for the reaction of steam methane reforming coupled with calcium looping, also known as sorption enhanced steam methane reforming (SE-SMR). In this study, the data was obtained using the Aspen Plus software, where input parameters, regenerator and reformer temperatures, pressure, steam-to-carbon ratio and sorbent-to-carbon ratio, were varied to obtain a wide range of data for the process. The Aspen Plus data was validated against literature data, and was then split into training, validation and test data. Various gas concentrations in the reformer and regenerator, as well as methane conversion were used as the output parameters. The models developed showed high accuracy prediction for the reactor gas concentrations and confirmed that ANN and RF algorithms can successfully model a nonlinear process such as SE-SMR, and therefore act as a suitable data-driven soft sensor for the process.

Krzywanski *et al.*[155] explored a method of predicting the NO_x emissions produced from the regenerator of a calcium looping system, coupled with oxyfuel combustion of coal to provide heat of decomposition, using a regression analysis-based modelling technique. The authors conducted the experiment in a dual-fluidised bed (DFB), with the effects of fuel type, oxygen feed, and NO addition to primary or secondary feed gas, being evaluated. The authors provided limited detail on the regression model, however Fig. 8 shows the flowchart of the model application, and the only input necessary are the fixed carbon, the ratio of molar nitrogen to carbon content in fuel N/C, and the O_2, concentration in the flue gas from the regenerator, leading to the NO_x emission as the output parameter. The results obtained from the model were in good agreement with experimental results, with a correlation coefficient equal to 0.925.

An alternate purpose that ML has been applied to in the calcium looping field, is on the study of the economic feasibility of the post-combustion calcium looping process on a 580 MW coal fired power plant, by Hanak and Manovic.[156] In this study,

Fig. 8 Application of the model for the evaluation of NO_x concentration if flue gas.[155]

an ANN was developed using data from Aspen Plus simulations, and this model was then combined with results from an economic model developed from a Monte Carlo (MC) simulation. The ANN model was used to connect the process inputs of the process model with the process inputs of the economic model. A two-layer feedforward ANN with ten sigmoid hidden neurons and linear output neurons was developed, with 70% of the data obtained from the Aspen Plus model, used for training, 15% used for validation and 15% used for testing. Fig. 9 shows that the ANN used in this study can depict the thermodynamic performance of the calcium looping retrofit accurately, despite its nonlinear characteristic. The study concluded that the stochastic approach, and incorporation of the ANN model, in the economic feasibility assessment enables a more accurate and reliable comparison of different calcium looping retrofit configurations.

3.3.3 Perspectives and prospects. ML has been successfully applied in oxy-fuel combustion for the combustion characteristics prediction and process monitoring. It should be pointed out that most researchers use TGA data to train, validate and test the ML models to predict combustion characteristics, but these also can be easily measured by the TGA without using the training data to develop the optimal ML model that requires higher computing costs and longer time. In addition, the extracted TGA data cannot represent the combustion characteristics in the real combustor due to their low heating rates and mass-heat transfer considerations. Thus, it is suggested that the researchers could use the data from the pilot-scale or large-scale combustors to develop their ML models, and the trained ML models could provide more useful information to develop oxyfuel combustion technology. ML can also be applied for using the flame images to monitor oxyfuel combustion process.

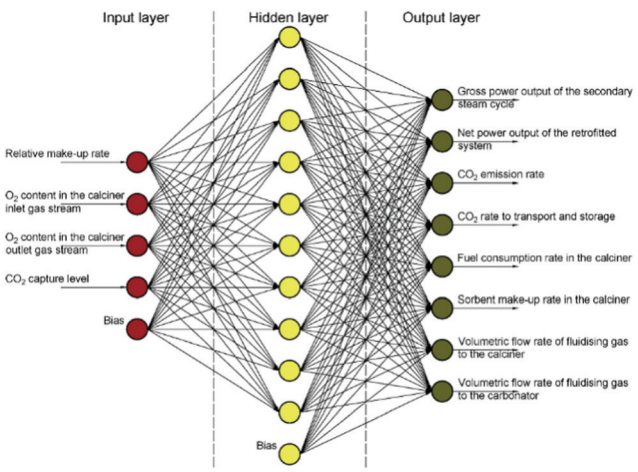


Fig. 9 Structure of the artificial neural network used to map the thermodynamic performance of the calcium looping process retrofit.[156]

For calcium and chemical looping technologies, it is expected that ML will play an important role in materials development, process control, and techno-economical assessment. However, only a few researchers have attempted to utilise ML for these goals. We encourage researchers working in this area to consider applying ML in their research to maximise their research outputs. For instance, CLC is a novel carbon capture technology, and the selection of suitable oxygen carriers is a key barrier to chemical looping technologies development. Over the last 20 years, over 1000 materials have been investigated experimentally. This could serve as an ideal database for utilising ML to screen and identify useful information to guide the oxygen-carrier materials development. Also, ML can be combined with density functional theory (DFT) to screen the thermodynamic feasible metal oxides as the oxygen carriers.[157] It is also foreseen that ML will accelerate the discovery, design, and synthesis of sorbents for calcium looping process by using the historical research data on sorbents development.

In the Section 3, we have reviewed and discussed the research of applying ML in CO_2 capture, which includes CO_2 absorption, CO_2 adsorption, oxyfuel combustion, calcium looping and chemical looping combustion. There is also work on ML in membrane for CO_2 separation which is detailed elsewhere.[158–160]

4. Machine learning in CO_2 transportation, utilisation, and storage

4.1 Machine learning in CO_2 transportation

4.1.1 The role of machine learning in the mass flow metering of CO_2. The captured CO_2 needs to be transported from the capture points to the storage sites. Pipeline transportation of CO_2 in the dense phase is regarded as the most cost-efficient and safest solution over a long distance.[161] Accurate flow metering of CO_2 in CCUS pipe networks is crucial to the optimised design and economical operation of CCUS processes. For instance, it is reported that each percent of accuracy improvement will save €200k per year for a CCUS project in Norway.[162] As expected, larger-scale CCUS systems, a higher number of accurate flowmeters need to be deployed. In addition, the European Union Emission Trading Scheme (EU-ETS) requires the flowmeters to operate within an uncertainty of ±1.5%.[161]

However, it is difficult for traditional flowmeters to meet the accuracy requirements due to the complex properties of CO_2 fluid. Unlike water, oil and natural gas, CO_2 is expected to be transported near the critical point, which is very close to the expected operational condition of transportation pipelines. A small change in line temperature and pressure may lead to a significant change in the phase of CO_2, resulting in gas–liquid two-phase CO_2 flow. Impurities produced using different capture methods may also affect the phase behaviours of CO_2 flow. In addition, some impurities, such as water, H_2S, NO and SO_2, produce corrosive products which may influence the choice of

flowmeter material.[163,164] For some volumetric flowmeters, the density data calculated from the equation of state (EoS) is required to obtain the mass flowrate. However, the accuracy of EoS of CO_2 flow with impurities is insufficient.[165] Moreover, flexible operations of CCUS systems on smart fossil fuel fired power plants, such as frequent load changes and rapid start-ups and shutdowns, may lead to rapid changes in the properties of CO_2 flow. Transient behaviours that occur in pipelines may result in the phase transition of CO_2 and flow instability, making the accurate measurement of CO_2 flowrate more challenging.

Over the past few decades, some techniques have been developed to achieve the accurate measurement of multiphase flow, especially gas–liquid two-phase flow. Some of these techniques, such as radiation attenuation and nuclear magnetic resonance, exhibit satisfactory performance in terms of measurement range and accuracy, and can directly provide mass flowrate, density and composition of multiphase flow.[166,167] Nevertheless, the high cost and system complexity restrict their applicability in the CCUS sector. Other economical techniques such as differential pressure-based flowmeters are not able to achieve satisfactory accuracy in the mass flow measurement. In order to improve the accuracy of flowmeters, low-cost sensing techniques incorporating ML algorithms have been proposed in recent years.[168,169] ML algorithms are capable of handing the hidden relationships in large, complex and multivariate datasets and have been used in the measurement of gas–liquid two-phase CO_2 flow.

4.1.2 Measurement of the mass flowrate of two-phase CO_2 flow. Mass flowrate measurement of CO_2 flow is essential for the fiscal purpose in CCUS projects. Coriolis mass flowmeters, as the most accurate single-phase mass flowmeters, have the ability of directly measuring mass flowrate, but the errors in measuring two-phase flow are still large. Thus, ML algorithms are employed to improve the accuracy of Coriolis mass flowmeters in multiphase flow measurement, based entirely on internally observed parameters. Fig. 10 shows the common solution based on Coriolis mass flowmeter and ML algorithms. The ML algorithms use input variables reading from Coriolis flowmeters and give the measured mass flowrate, density, and gas volume fraction (GVF). When CO_2 flow is single-phase liquid or gas, the output of GVF is 0% and 100%, respectively.

Henry *et al.*[171] reported a case study which achieved the errors of mass flowrate within 1–5% in the measurement of gas–oil two-phase flow based on a Coriolis mass flowmeter and an ANN under the condition of 1 kg s^{-1} to 10 kg s^{-1} in flowrate and less than 60% in GVF. The same measurement system was also employed to measure slugging two-phase CO_2 flow at the pressure of 5.52–7.03 MPa and the temperature of 4–32 °C.[172] Results show that the reading difference between the Coriolis flowmeter and other sales meters over several weeks is usually within ±5%. Comparative investigations into the performance of ML algorithms for gas–water two-phase flow metering were conducted by Wang *et al.*[173] Several algorithms, such as ANNs, SVM and GP, were developed to estimate the liquid mass flowrate and GVF. The inputs of the ML algorithms were obtained from a Coriolis flowmeter and a differential pressure (DP) transducer. For the mass flowrate measurement, the input variables are apparent mass flowrate, apparent density, damping and DP, while for the GVF measurement, the apparent mass flowrate, density and DP are taken as inputs. Results show that the relative errors are within ±1% in mass flowrate measurement over the range of 250 to 3200 kg h^{-1} and within ±10% in GVF prediction. Wang *et al.*[170] also applied a Coriolis mass flowmeter incorporating LS-SVM models to measure the mass flowrate of gas–liquid two-phase CO_2 flow in both horizontal and vertical pipelines. Fig. 11 illustrates the principle of the flow measurement of gas–liquid two-phase CO_2 flow. A classification model is developed and incorporated in the system to recognise the flow pattern and independent LS-SVM models for the mass flowrate metering of gas–liquid two-phase CO_2 flow. Results suggest that most of the relative errors under steady-state flow conditions are within ±2% in horizontal test pipeline and ±1.5% in vertical test pipeline. However, the performance of the models is affected by the lack of verification under dynamic flow conditions. It should be noted that the aforementioned models can also be trained to measure the GVF of two-phase CO_2 flow (Section 4.1.3).

4.1.3 Measurement of the gas volume fraction of two-phase CO_2 flow. Accurate GVF measurement of gas–liquid two-phase CO_2 flow in a pipeline network is crucial to the safe and economic operation of the CCUS process. In recent years, some accessible sensing solutions such as capacitive sensors

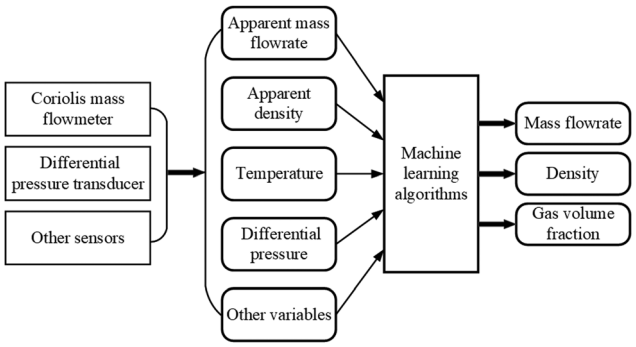


Fig. 10 A typical CO_2 flow measurement system based on low-cost sensors and ML algorithms.[170]

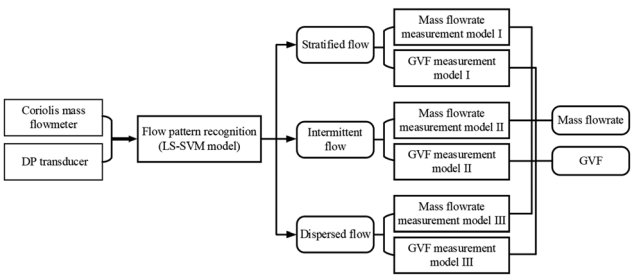


Fig. 11 Principle of the mass flowrate and GVF measurements of two-phase CO_2 flow.[170]

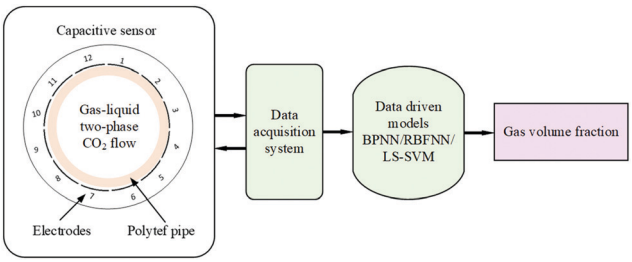


Fig. 12 Principle of CO_2 GVF measurement using capacitive sensors.[27]

and Coriolis flowmeters in conjunction with ML algorithms have been proposed to measure the GVF of CO_2 flow.

As shown in Fig. 12, a flow-pattern-based LS-SVM model developed by Wang *et al.*[173] was utilised to measure the GVF of gas–liquid two-phase CO_2 flow. Experimental results suggest that errors of the measured GVF are mostly within ±10%. Shao *et al.*[27] achieved the GVF measurement in a horizontal CO_2 pipeline based on a 12-electrode capacitive sensor and data-driven models, as shown in Fig. 12. Three data-driven models, BPNN, RBFNN and LS-SVM, were established. Unlike the flow pattern recognition approach, reconstructed images are usually not required for GVF measurement. The GVF measurement of two-phase CO_2 flow is achieved without the time-consuming image reconstruction of the flow pattern. Experiments were conducted under both steady-state and dynamic flow conditions. For steady-state flow conditions, the mass flowrate was set from 200 to 3100 kg h^{-1} while the GVF was from 0–84%. Under dynamic flow conditions the gas phase CO_2 was rapidly increased from 120 kg h^{-1} to 400 kg h^{-1} and then decreased while the liquid CO_2 was fixed at 1500 kg h^{-1}. Measurement results show that the RBFNN outperforms the other two models. Errors are mostly within ±7% and ±16% under steady-state and dynamic flow conditions, respectively.

4.1.4 Input variable selection for CO_2 flow metering

Significance of variable selection in ML. Input variable selection is an essential step in the development of ML models. It is intended to eliminate the irrelevant or redundant variables from the available data, which is directly obtained from sensors or in a transformed manner and identify a suitable subset which is significant to estimation of the desired output. Due to the inherent complexity of multiphase flow and the limited theoretical knowledge of complex physical phenomena, input variable selection becomes more important. Input variable selection is helpful to analyse parametric dependency between input variables and their significance and sensitivity to the desired model output. Meanwhile, it is beneficial to reduce the complexity of the model structure and improve the computational efficiency of the model. Therefore, input variable selection should be considered before developing ML models.

It must be pointed out that dimension reduction algorithms such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are easily confused with input variable selection. Dimension reduction aims to transform data from a high-dimensional space into a low-dimensional space, resulting in a reduced number of variables.

Methods that may be used to select variables. Input variable selection techniques can be classified into three main categories: wrapper, embedded and filter algorithms. Wrapper algorithms, such as Single Variable Regression and Genetic Algorithm-Artificial Neural Network (GA-ANN), and embedded algorithms, including Recursive Feature Elimination and Evolutionary ANNs, are model-based, *i.e.*, a model has to be constructed and trained in advance. Filter algorithms such as Rank Correlation, Partial Correlation and Partial Mutual Information (PMI) are model-free. May *et al.*[174] considered several key factors in determining the most appropriate approach to input variable selection for a given application. The model-based approach aims to select the variable set which makes the model perform well through establishing and evaluating the model through potential variable combinations. The main drawback of this approach is the high computational requirement due to a large number of calibration and validation processes required. Moreover, the selection results depend on the predefined model in terms of architecture and parameters. By contrast, the model-free approach is directly based on the information (interclass distance, statistical dependence, or information theory, *etc.*) between the available dataset, so the computational efficiency is not an issue. However, a trade-off criterion should be defined to balance the significance measurement and the number of selected variables.

For air–water two-phase flow measurement, Wang *et al.*[173] applied PMI, GA-ANN and tree-based iterative input selection (IIS) methods to investigate the parametric dependence, significance and sensitivity of the input variables to the desired outputs, *i.e.*, mass flowrate and GVF. Results suggested that the selected variables using the PMI algorithm, observed density, apparent mass flowrate, DP and damping provide more effective information for the models to measure liquid mass flowrate. The variables selected using the tree-based IIS algorithm, included observed density, apparent mass flowrate and DP, which were more significant to predict GVF. Subsequently, Wang *et al.*[170] investigated the measurement of gas–liquid two-phase CO_2 flow and developed LS-SVM models for flow pattern recognition, mass flow measurement and GVF prediction (Section 4.1.3), with the selected input variables including apparent mass flowrate, observed density, damping and DP.

Although variable selection approaches can provide some valuable information to determine the input variables of an ML model, the accuracy of the methods also depend on the observational dataset, such as data size and their distributions. A dataset with less data or low-quality may result in underestimation or overestimation of the candidate variables for an ML model. Consequently, in order to ensure the selection accuracy with a limited size of a dataset, it is necessary to determine the input subset by combining variable selection methods with engineering judgement based on the relevant knowledge of the target application. The results of input variable selection will help enhance engineering judgement whilst the latter will interpret the variable selection results.

4.1.5 Perspectives and prospects

CO_2 flow metering under steady and dynamic conditions. Although CO_2 flow metering has achieved higher accuracies

under steady flow conditions, the online implementation and in-situ calibration of a data driven model should be incorporated. In addition, smart power plants with CCUS facilities are required to balance the power grid by compensating for the intermittent electricity supply from renewable energy resources such as wind farms and solar stations. As a result, smart CCUS plants will need to be operated flexibly.[175,176] Load change, frequent start-up and shutdown will occur during flexible CCUS operations, which will generate constantly occurring transient flow conditions. Recent experimental investigations revealed significant discrepancies in the mass flow rate of two-phase CO_2 between the measured value from a Coriolis flowmeter and the reference value during the load change in a CO_2 transportation pipeline, which could lead to significant errors in the fiscal metering of CO_2.[177] Therefore, CO_2 flow metering with a ML model that considers dynamic nature of the flow, such as a dynamic neural network should be investigated.

Deep learning algorithms of Long Short-Term Memory (LSTM) and Gate Recurrent Unit (GRU) may also offer possible solutions. Meanwhile, a data driven model is usually a black-box which is highly dependent on the available dataset and it may result in poor generalization capability when used on practical CCUS facilities. ML by combining a physical based model and a data driven model may improve the model interpretability, measurement accuracy and generalization capability, but further research is required in this direction. In addition, the data driven models that have been proposed and developed to date have some drawbacks, such as heavy computational workload caused by the feature engineering or inefficiency when dealing with a high volume of data. Therefore, the necessity and significance of developing new deep learning models, which can deal with the above problems, should be investigated.

Mass flowrate metering of CO_2 with impurities. There are a range of impurities such as N_2, Ar and O_2 in a CO_2 stream from fossil fuel power plants and large-scale industrial emitters. Such impurities have a potentially significant influence on the thermophysical properties of CO_2 and hence large errors in the mass flow metering of CO_2. In addition, the range and level of impurities in a CO_2 stream vary under different carbon capture sources.[178] As a result, the flow measurement system should combine the information from the mass flowrate and the GVF to obtain the actual mass flowrate of CO_2 component in the presence of impurities.

A reliable CO_2 test rig is essential for R&D in CO_2 mass flow metering of single-phase and two-phase CO_2 with impurities under both static and dynamic CCUS conditions. A dedicated CO_2 two-phase flow rig with an inner pipe diameter of 25 mm is available at the North China Electric Power University. The liquid flowrate of CO_2 ranges from 200 to 3600 kg h^{-1} with uncertainty of 0.16%, while the gas flowrate range is from 15 to 400 kg h^{-1} with uncertainty of 0.3%. The line pressure of the rig can be varied from 57 to 72 bar with a temperature between 20 and 30 °C. However, new features, including a wider range of flow conditions, injection of impurities, different pipe orientations for meters under test, and variations in the pipe diameter of the test sections should be developed in future.

Leakage detection of CO_2 from transportation pipelines and from storage sites. Potential CO_2 leakages from high-pressure CO_2 transportation pipelines and from storage sites pose a significant threat to the safety and health of those living in the vicinity of CCUS pipe networks and storage sites. The possibility that CO_2 may migrate from storage sites is a primary concern for the safety and effectiveness of the CCUS technologies. Permanent, automated monitoring techniques for the continuous leakage detection of CO_2 from transportation pipelines and storage sites are necessary. For the CO_2 leakage detection in transportation pipelines, although acoustic emission (AE) sensors have been applied to locate the position of the leakage source,[179] the flowrate of the CO_2 leakage needs to be estimated. By combining the information from the AE sensors and relevant temperature and pressure data, a leakage location and estimation model based on ML algorithms should be developed for the safe operation of the CCUS pipe networks. Moreover, for the large-area monitoring of a CO_2 storage site, remote sensing techniques, such as hyperspectral imaging, aerial imaging and satellite imaging, should be considered.[180–182] Meanwhile, in-field pressure and seismic transducers may also be applied for the local-area monitoring of a CO_2 storage site.[183] An integrated monitoring system by fusing the information from the remote imaging systems and from the in-field transducers is a promising solution, which should facilitate the practical deployment of CCUS technologies.

4.2 Machine learning in CO_2 storage and utilisation

4.2.1 CO_2 storage. Ideal CO_2 storage places include saline aquifers and depleted oil reservoirs because of their high storage capacity with available infrastructure[184] (*i.e.* caprocks that prevent the migration of CO_2 plume) in place. More importantly; the injection and production wells in those mature fields can serve as the injection path for CO_2 storage.[185–187]

Four types of trappings could occur when CO_2 is injected into depleted oil reservoirs: structural-stratigraphic trapping, solubility trapping, residual trapping, and mineral trappings.[188–190] Structural-stratigraphic trapping is the process that CO_2 is stored in the underground structure as a supercritical state.[191] CO_2 can often be trapped under low permeable formations such as shale or mudstone, which can prevent CO_2 from migrating upward due to the buoyancy force. Besides, impermeable zones such as cap rocks and sealed faults can also provide a good condition for the entrapment of CO_2.[192,193] Thus, the investigation of the caprock integrity for a long-term sealing capability is important before a CO_2 sequestration project is carried on.[194] Solubility trapping refers to the dissolution of CO_2 in the formation of aqueous and oleic phases.[195] The solubility of CO_2 in formation water depends on underground conditions including pressure, temperature and water salinity. Numerous studies have been performed to construct the relations between the CO_2 solubility with those parameters that would impact solubility trapping (*i.e.* diffusivity,[196] oil/gas–brine interfacial tension (IFT),[197] *etc.*).

The solubility of CO_2 in the oil phase is generally higher than that of brine in mature oil reservoirs.[191] Residual trapping involves the process that trapping CO_2 as an immobile phase within the porous media due to capillary forces. It is an important phenomenon in the CO_2 sequestration process especially when there are no reliable sealing formations or caprock. The gas hysteresis effect plays a vital role in the residual trapping.[198] The bypass of a wetting phase fluid will render the non-wetting phase immobile, thus leading to the entrapment of the non-wetting phase. The effect of residual trapping can be enhanced when the hysteresis effect is considered. Ampomah *et al.*[191] in a detailed numerical simulation study, pointed out that there would be an apparent increase in the predicted amount of CO_2 trapped as a residual phase after the gas hysteresis effect was implemented. The predicted residual trapped CO_2 surged from 1% to 14% after the hysteresis effect was considered. In the mineral trapping, CO_2 will react with formation mineralogy and be trapped in the precipitation or dissolution of extant or new carbonate minerals. Compared with other mechanisms, CO_2 reactions often take years to occur thus its impact on the transportation of the CO_2 plume would be observed on a longer time scale. When CO_2 is in contact with formation brine, aqueous species such as soluble CO_2, HCO_3^-, CO_3^- are generated, and then reacted with formation minerals. Some common reactions between CO_2 and formation mineralogy are summarised in Table 4.

Several studies using ML-based methodologies have been performed regarding how those trapping mechanisms influence the dispersal and migration of the CO_2 plume. Sun *et al.*[188] studied the CO_2 trapping mechanisms in the Morrow B Sandstone in the Farnsworth Units. A neural network-based approach was used to match the reservoir model with historical data. The history matched model was then employed to evaluate the impacts of residual, structural-stratigraphic, solubility, and mineral trapping mechanisms on CO_2 sequestration and hydrocarbon production. The ML-based history match process was able to provide reliable pressure, fluid saturation and composition distributions that help the numerical model effectively investigate trapping mechanisms with a reduced computational overhead. The conclusion was that more CO_2 is dissolved in the oleic phase than the aqueous phase, which is due to the high salinity of the formation water. Moreover, mineral trapping plays a less significant role in the CO_2 sequestration process compared with other trapping mechanisms.

Ni and Benson[199] studied the effect of mesoscale heterogeneity on larger-scale multiphase fluid flow properties and trapping behaviours using a ML clustering method. The CO_2 saturation maps, the voxel-level porosity and the permeability maps were used as the inputs for the model. Each voxel was treated as one data point, and the time series properties at each voxel were treated as individual attributes (*i.e.*, CO_2 saturation time series). The CO_2 saturation and the porosity maps were obtained through CT image manipulation, and the voxel-level permeability map was obtained using the extended Krause's method.[199] This study tested two clustering methods and found that *K*-means clustering was more suitable for characterizing flow behaviours and hierarchical clustering was more desirable for identifying the capillary heterogeneity trapping behaviours. Five different sets of coreflooding data were used to examine the feasibility of the proposed approach. They concluded this method was able to assess how the mesoscale petrophysical properties influence capillary-dominated flow and residual trapping behaviours. Moreover, the differences in time series behaviours among the different clusters would be diminished in viscous-dominated flow regimes.

CO_2 storage of solubility trapping involves the process where the injected CO_2 contacts *in situ* brines and dissolves into the water through molecular diffusion. Research was carried out to study the CO_2/oil/brine interactions under subsurface conditions. Amar and Ghahfarokhi[196] established the correlation between diffusivity coefficients of the CO_2 in brine water with pressure, temperature and the viscosity of the solvent using the group method of data handling (GMDH) and gene expression programming (GEP). GMDH is one type of ANN that can generate an explicit expression for the correlation between inputs and output. The correlation generated using GMDH takes the advantage of polynomial models. GEP is one evolutionary technique to mimic systems with accurate explicit expressions, which is an improved version of genetic programming. Besides the common genetic operators, including selection, crossover, elitism and mutation, GEP also introduces new actions such as insertion and transposition to find a reliable correlation. The conclusion was that both GEP and GMDH correlations were able to make predictions that were very close to experimental values, and the GEP correlation yielded higher accuracy than the GMDH correlation. The GEP model was also compared with decision trees (DTs), RF, mixed Kernel-based SVM coupled with GA and other pre-existing models, the GEP model was superior to all these models.

Menad *et al.*[200] proposed to use MLP and RBFNN to predict the CO_2 solubility in brine at different temperatures, pressures and molalities of NaCl. Additionally, several evolutionary algorithms were employed to optimise the control parameters of the neural networks, namely the Levenberg–Marquardt (LM) algorithm, GA, particle swarm optimization (PSO) and artificial bee colony (ABC). Combinations of those methods were compared to determine the best one. They found that RBFNN-ABC

Table 4 Summary of some common reactions between CO_2 and formation mineralogy[188–190]

	Reactions
1	$CO_{2(aq)} + H_2O = H^+ + HCO_3^-$
2	$CO_3^{2-} + H^+ = HCO_3^-$
3	$OH^- + H^+ = H_2O$
4	Quartz = $SiO_{2(aq)}$
5	Albite + $4H^+ = 2H_2O + Na^+ + Al_3^+ + 3SiO_{2(aq)}$
6	Calcite + $H^+ = Ca^{2+} + HCO_3^-$
7	Dolomite + $2H^+ = Ca^{2+} + Mg^{2+} + 2HCO_3^-$
8	Siderite + $H^+ = HCO_{3-} + Fe^{2+}$
9	Illite + $8H^+ = 5H_2O + 0.6K^+ + 0.25Mg^{2+} + 2.3Al^{3+} + 3.5SiO_{2(aq)}$
10	Kaolinite + $6H^+ = 5H_2O + 2Al^{3+} + 2SiO_{2(aq)}$
11	Smectite-low-Fe–Mg + $7H^+ = 0.29Fe^{2+} + 3.75SiO_{2(aq)} + 0.16Fe^{3+} + 4.5H_2O + 1.25Al^{3+} + 0.15Na^+ + 0.02Ca^{2+} + 0.2K^+ + 0.9Mg^{2+}$
12	Chamosite-7A + $10H^+ = 2Fe^{2+} + SiO_{2(aq)} + 2Al^{3+} + 7H_2O$

would yield to the most accurate prediction in the tests among all combinations.

Zhang *et al.*[201] proposed a work to model the CO_2–brine IFT using extreme gradient boosting (XGBoost) trees. The generated model was then employed to determine the optimal CO_2 sequestration depth in saline aquifers. The brines used to synthesise the database consider one or more of the following salts: NaCl, KCl, Na_2SO_4, $MgCl_2$, and $CaCl_2$. Thus, the total molalities of the monovalent cations (Na^+ and K^+) and bivalent cations (Ca^{2+} and Mg^{2+}) were considered as two independent input variables. CH_4 or N_2 were two impurities accounted for in the CO_2 stream, so the mole fractions of these two impure components were categorised into other two individual input variables. Pressure and temperature were also utilised as the other two variables due to their important impacts on the CO_2–brine IFT. After inconsistent data points were removed, a total of 2346 data points were used to train the IFT prediction model. The XGBoost trees model combined a cluster of classification and regression trees (CARTs) to fit the training data samples. The basic components contained in CART are a root node, a set of internal nodes, and a set of leaf nodes, which is depicted in Fig. 13.

The hyperparameters of the XGBoost trees were optimised using the K-fold cross-validation integrated with the exhaustive grid search approach. In the grid search approach, the search range of each parameter is divided into different grids and this approach will test the values of all grids to determine the best result. Based on the model, the permutation importance (PI) was employed to ascertain the importance of each input variable to the IFT. Results showed that pressure had the highest impact on IFT, followed by temperature, bivalent cation molality and monovalent cation molality, while the mole fractions of CH_4 or N_2 were the least important factors. The capacity of structural trapping CO_2 in aquifers varies with the CO_2–brine IFT that would be affected with different temperatures and pressures. It was claimed that with the help of the generated model, reservoirs with different pressure and geothermal gradients can be used to study the capacity of structural trapping CO_2. An increase in the maximal structural trapping capacities for shallower formations was observed when the pressure was higher and/or the geothermal gradient was lower.

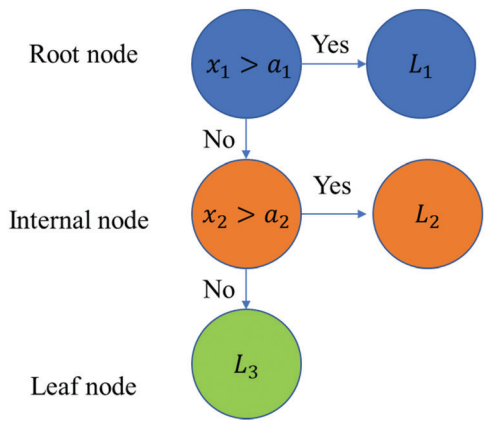


Fig. 13 Illustration of a CART. L denotes the leaf node (modified from Zhang *et al.*[201]).

CO_2 leakage detection. After the CO_2 is injected into the subsurface complex, it is necessary to use monitoring and verification approaches to ensure the safe and long-term storage of injected CO_2.[202] The common method includes building a numerical model to simulate how the CO_2 plume moves in the underground structure and to predict the feasibility of the long-term storage of the sequestered CO_2.[203] Direct or non-direct monitor data is always utilised in collaboration with numerical models to assess risk of CO_2 plume leaks from faults, legacy well, or fracture systems.[204]

Wang *et al.*[205] studied how to interpret the CO_2 saturation using seismic and downhole monitoring data. This study used ML approaches to infer the CO_2 saturation at different depths from the combination of synthetic seismic data and monitored downhole pressure and total dissolved solids (TDS) information. The framework was built upon a candidate geologic carbon storage site near Kimberlina, CA, USA. A hypothetical well leakage was included in the numerical model, which was focused on simulating the three geological layers overlying the CO_2 storage reservoir. All three layers were aquifer layers with a sand fraction of approximately 0.8. There were 6000 numerical simulations implemented by varying the distributions for the permeability of the three geologic layers. Each simulation had a 20 years' prediction with a timestep of one year. At each time step, rock physics modelling was performed to estimate changes in seismic velocity due to the simulated CO_2 and brine leakage from the flow simulation outputs. Therefore, a total of 120 000 forward seismic velocity models were obtained from those 6000 simulations. Each velocity model was further used to generate synthetic shot gathers using 2D finite-difference acoustic wave modelling, along a sparse 2D seismic line with only five shots and 40 receivers. For each velocity model, five seismic features were calculated thus 1200 ($= 6 \times 40 \times 5$) seismic features could be used to train the prediction model. Besides the seismic features, measured downhole pressure and TDS at three depths were also included in the training inputs, leading to a total of 1206 involved in each input-output pair. The output was the category of CO_2 saturation at three depths that have been labelled as five different integers to discretize the range of CO_2 saturation from zero to very high level. The SVM with a linear kernel (linear SVM), support vector machine with a radial basis kernel (SVMr), DNN with two hidden layers and recurrent neural network (RNN) with a LSTM layer were used to train the CO_2 saturation prediction model respectively. The performance of the models was estimated using the Kappa statistic, meaning the prediction accuracy was calculated and ranked between 0 to 1, with 0 representing a random prediction and 1 standing for perfect prediction. It was concluded that compared with using seismic monitoring alone, adding downhole pressure and TDS measurements as input features could improve the accuracy of the CO_2 saturation inversion.

Sinha *et al.*[28,183] demonstrated how to detect the CO_2 leakage using pressure data. The injection of CO_2 would cause pressure perturbation across the reservoir field. Harmonic pulse testing (HPT) is one approach to cause this kind of perturbation hence it can be used to differentiate CO_2 leakage. In a typical HPT job, the perturbation was induced by the harmonic injection of a fluid into the reservoir at the injection well, and the responses were recorded at the observation well. The pressure HPT can be used to differentiate the pressure response of a leak *versus* the non-leak in a field test. In a CCUS project across multiple depleted oil fields, many injection wells and abandoned wells could act as the path for CO_2 leakage, making the interpretation of the voluminous HPT data a challenging task for human brains. However, the ML techniques can be a good alternative. In this work, the author used different neural networks to build the anomaly detectors to interpret CO_2 leakage, including multi-layer neural network (MFNN), LSTM, convolutional, neural networks (CNN), and a combination of CNN and LSTM (CONV-LSTM). The actual measured pressure signal was compared with the predicted response for the non-leak situation, and then the error was calculated as an indicator of the CO_2 leakage (anomaly). The conclusion was that LSTM outperformed the others in the pressure anomaly detection tests and the proposed approach could provide early warnings to the CO_2 leakage in a CCUS project.

Lima and Lin[206] integrated geological data and ML techniques to predict the CO_2 and brine leakage in a 200 years' duration in geological carbon sequestration (GCS) project. The database used for the employed machine-learning approaches was acquired from 500 simulations that were generated to model underground water flow and understanding effects at GCS sites attributed to CO_2 injection. Those models contain an injection well, a legacy well and three geological layers. The seismic data and legacy well pressure was used as inputs for function predicting CO_2 and brine leakage amount. The Inception model was used to train the seismic data and CNN model was used to handle pressure data. Here, 50 out of 500 simulations were utilised as test sets, and models' performance was compared between the model only using seismic data and other using both seismic data and well pressure. It was found that including pressure data would provide small improvements in the prediction of CO_2 and brine leakage. Moreover, employing this developed approach was able to provide an accurate prediction of the CO_2 and brine leakage on GCS sites.

Zhong *et al.*[207] used a combined CNN and LSTM model, designated as ConvLSTM, to detect the CO_2 leakage in a CCUS project. The CNN model was used to handle the spatial features and the LSTM was used for temporal features. The spatial features considered porosity and permeability and the temporal features included the CO_2 injection rate and the bottomhole pressures of a production well and a leak well. The temporal features were transferred into 2D images and the pixel value at the injection well location was the injection rate and the pixel values at the production and monitor wells were corresponding bottomhole pressures. Thus, the total inputs for the ConvLSTM model were three 2D images including one image containing the injection rate and bottomhole pressure at the production well, and the other two are areal distributions of the porosity and permeability. The output from the model was the predicted bottomhole pressure at the monitoring well, which was compared with a real monitored pressure to determine whether there is an anomaly in the CO_2 injection. The database used to train the ConvLSTM model was from a pulse testing experiment where the CO_2 is injected cyclically with an injection duration of 90 minutes. The injected CO_2 was artificially produced at a constant production rate of 60 kg min^{-1} to mimic a CO_2 leakage at the production well. A detection function was defined to calculate the probability of the test data point being in a user-defined normal data range given a user-defined threshold. They also pointed out that insufficient datasets or existing noises in the raw data may lead to inaccurate prediction.

Singh[208] introduced a workflow to monitor and detect CO_2 leakage from a reservoir using injection rates and bottomhole pressures. A deconvolution response was defined as the function of time-dependent well bottomhole pressure and injection rates to measure the fluid leakage, which could be simulated using MLR of all the wells present in the reservoir. The model training process followed a strategy that field history without any leakage was used to train and validate the model. Then the model prediction was the simulated scenarios where no leakage took place. The deviation between the predictions and real monitoring deconvolution responses was employed to determine the leakage. The capability of the proposed workflow was demonstrated by applying it to three case studies: (1) a naturally fractured tight reservoir with five injectors and four monitoring wells; (2) a reservoir with a barrier and the same well pattern as case 1; (3) a real deep offshore saline aquifer with thick shale layer above and below the reservoir. It was concluded that the proposed method was able to detect leakage of both incompressible and compressible fluids from a simple reservoir to a fully heterogeneous and structurally complex field. The author also pointed that this method could provide preliminary insights into the location of the leakage, but still required the help of expensive surveys (such as seismic, *etc.*) to identify the actual location of a leak and the severity of the leak.

4.2.2 CO_2 utilisation

4.2.2.1 CO_2-Enhanced oil recovery. The utilisation of CO_2 as an injecting phase for enhanced oil recovery (EOR) has decades' of history.[209–211] CO_2-EOR is a widely used technique that injecting CO_2 into a reservoir after waterflooding to lower the residual oil saturation and hence improving hydrocarbon production.[212–215] When the injection CO_2 enters the subsurface, a large volume of the injected CO_2 will be trapped underground due to the effects of the aforementioned trapping mechanisms.[216] Thus, the applications of CO_2-EOR with CCS would have dual benefits that both extracting more oil and injecting and sequestering anthropogenic CO_2.[217,218]

The applications of ML-based approaches mostly seek to reduce the computational overhead required by calling for the original high-fidelity numerical model,[219,220] hence shortening

the time needed by running the numerical model and further enabling some complicated jobs such as optimisation,[221,222] and uncertainty assessment.[214] This type of application is often considered as generating a proxy model or surrogate model using various ML-based approaches.

Vida *et al.*[223] introduced a work that couples grid-based surrogate reservoir model (SRM_G) and well-based surrogate reservoir model (SRM_W) to simulate a CO_2-EOR project at the Scurry Area Canyon Reef Operators Committee (SACROC) oilfield. The SRM_W models were used to investigate the flooding front and simulate the changes in properties along with time in each grid block in the reservoir. The properties that were handled by SRM_G included pressure, phase saturation, or composition of reservoir fluid components at any desired time step. The SRM_Ws were used to deal with simulation related to well production data, such as oil rate, water rate and water oil ratio, *etc.* SRM_Ws could be used to estimate response of the reservoir at the well level (rate) to various reservoir parameters or operational constraints. An ANN model with one hidden layer was used to train the SRMs. The values of each property at each timestep were predicted using one trained SRM. For the SRM_G, a total of 60 neural networks were generated to predict the interested properties at each timestep (15 models per property). The integration of the SRM_Gs and SRM_Ws contained the following steps: at the initial timestep, SRM_Gs ran first and the calculated pressure, phase saturation, and CO_2 mole fraction for all grids were processed to obtain the well productivity index and tiering computations pertaining to grid-based and well-based systems. The information along with well-based initial information was then fed to SRM_Ws to calculate water, oil and CO_2 production at each well and entire field at first timestep. This process then proceeded to next timestep and information of each grid was updated until final timestep was reached. It was reported that total time for running 60 neural network models to deploy the SRMs' calculation was around 800 seconds. The original numerical model took more than 48 hours to run one realization that was used for optimization design on a machine with 24 GB RAM and a 3.47 GHz processor. By using coupled SRM models, one simulation job was finished in 15 seconds on the same computer.

Artun[224] studied single-well cyclic gas (N_2, CO_2 and CH_4) injection in fractured and depleted reservoirs. Various simulation scenarios were conducted based upon compositional reservoir model with hydraulically fractured well and low-permeable formations. This study focused on assessing impacts of design parameters on both volumetric and economic utilisation efficiency factors. Factors considered included the injection rate, duration (and volume), soaking duration, economic rate limit, and injected gas composition. A fast economic efficiency indicator was also constructed using neural networks based on the prepared simulating data. It was concluded that N_2 was better than other gases for short-term (5 or 10 years) benefits. Amini *et al.*[225,226] used SRM_G to replace the numerical reservoir model of a field located in Otway Basin in Australia with a CO_2 sequestration pilot project. The SRM model was trained through neural networks that used well data, static data and dynamic data as training inputs. It was concluded that the developed SRM model could generate outputs of complex reservoir models with high accuracy in a short time.

Amini and Mohaghegh[227] proposed work to develop proxy fluid flow model for the reservoir responses (pressure, saturation, and CO_2 mole fraction) undergoing a CO_2 sequestration process. The proposed approach was applied to a heterogeneous reservoir with 100 000 active grid blocks to verify its capability. During the reservoir simulation, properties at a certain grid block would depend on its interactions with the surrounding grids. For instance, the CO_2 movement and gas saturation at one grid would be affected by the pore volumes and degree of tightness of the grids in the vicinity of this grid. To account for this kind of dependence, tier systems were introduced to express the relationship between one specific grid to its surrounding grids. An ANN-based SRM model was generated using the data gathered from a CO_2 injection reservoir with one injector and one producer. Five different simulating scenarios were prepared by varying the CO_2 injection rates and cumulative injection volume. The training inputs included static data (grid location, grid top, porosity, permeability), calculated static data (distance to the injection well, distance to the sealing and non-sealing boundaries, user-defined parameters), well data (injection rate, cumulative injection) and the average porosities and permeabilities of the tier system; the training outputs were the dynamic data (pressure, gas saturation and CO_2 mole fraction at any timestep). An ANN model with one hidden layer was used to train the proxy. It was concluded that the computational speed was increased by about 20 times for this specific simulation case with an acceptable error margin.

Besides boosting computational speed, another reason for the employment of ML techniques is to ease the complexity of solving a problem, figuring out the unclear input–output patterns and structures that exist in the obtained experimental/simulated database. This mostly occurs when traditional methods fail to work properly due to missing information. As one of the critical parameters considered in the CO_2 flooding process, the precise prediction of minimum miscibility pressure (MMP) of oil in the CO_2-EOR process are widely studied. Sinha *et al.*[28] used ML techniques to predict MMP. The proposed method included using an analytical correlation that employed the SVM to tune the coefficients and a hybrid method that combined RF regression and generated correlation. A correlation was used to predict the MMP and linear SVM was used to tune the coefficients included in this correlation. It was reported the proposed correlation would work for spectrum of MMP from 6 to 34 MPa.

Xiong *et al.*[228] used two different methods to forecast unconventional reservoir well production, namely ANN and Time Series Analysis. Traditional methods such as decline curve analysis may not be as powerful as they normally would be when dealing with conventional reservoir well production due to limitations with shale oil production such as boundary dominated flow and constant operation condition. Peak production rate and hydraulic fracture parameters were considered as factors influencing oil production. DNN and autoregressive

integrated moving average (ARIMA) models were employed for the study. The ARIMA models updated their training data as function of time, thus a smaller time step will lead to more accurate predictions compared with real data. Moosavi *et al.*[229] tested the capability of four different hybrid-RBF networks in predicting oil recovery factor and oil rate in a foam-CO_2 flooding reservoir. The RBF network was combined with various evolutionary algorithms, namely particle swarm, imperialist competitive, genetic and teaching–learning based algorithm, to build the prediction model. These algorithms were employed to optimise the values for the weights and biases applied to the network nodes. It was claimed that teaching–learning-based optimization hybrid model (TLBO-RBF) achieved the greatest accuracy in predicting based on the datasets used in this study.

Chen *et al.*[230] developed a work to characterise the CO_2-EOR in residual oil zones (ROZ). ROZs are aquifers (or parts of aquifers) in which oil has migrated from source rock but is subsequently swept by the natural movement of aquifer waters over geologic time and remains at residual saturation. The main distinction between CO_2 storage in ROZs and conventional oil reservoir and brine was also assessed. Here, a ML models to predict potential of hydrocarbon production and CO_2 sequestration amount in ROZs were developed. Three ML models, namely Multivariate Adaptive Regression Splines (MARS), SVR and RF, were used and compared in terms of predictive capability in this work. It was concluded that when crude oil was present, more CO_2 would be dissolved in oil than brine water; while when there was no oil within the system, more gas would be trapped in the pore structure than be dissolved in the aquifer.

4.2.2.2 Optimising CO_2-CCS-EOR and uncertainty assessment. The utilization of ML algorithms in CO_2-CCS-EOR is often accompanied by optimization and uncertainty assessment work, in which a large volume of computations is needed. The ML model can be applied to generate proxy models as alternative to numerical model and reducing total computational time. Sun[231] employed a deep reinforcement learning method, namely the deep Q-learning (DQL) algorithm, to handle optimization of carbon storage reservoir management. The problem was treated as a Markov Decision Process (MDP), which was to model the intelligent agent's sequential interactions with an environment to obtain maximal returns. The key procedure of solving a MDP was to find the optimal value of the state-action function (Q-function) to have the best reward at each state without concerns about future states.[231] In DQL, the deep Q network (DQN) was used to approximate Q-function for quick investigation and response. Another target network was used to calculate the rewards at future states. To speed up the evaluation of a large number of system transitions by using DQL, a DL-based surrogate model was built up to accelerate the policy search process. The deep multi-task learning (deepMTL) was utilised to reflect correlations between pressure/saturation and selected inputs. A U-shaped architecture employing CNN as the building block was adopted to facilitate prediction of saturation and pressure simultaneously.

Menad and Noureddine[232] introduced a methodology to optimise CO_2 water-alternating-gas (CO_2-WAG) processes using NSGA-II (Non-Dominated Sorting Genetic Algorithm version II) coupled with a hybrid model based on MLP. LM, Bayesian Regularization (BR) and scaled conjugate gradient (SCG) algorithms were utilised in training proxy model. The objectives of this work were to optimise total oil recovery and total field water production. A total of 75 simulation realizations were generated using Latin Hyper Cube method and then fed to train a proxy model. The author concluded that the MLP-LMA model was the most accurate proxy. Zhang and Sahinidis[233] employed polynomial chaos expansion (PCE) to generate a proxy model used in uncertainty quantification in CO_2 sequestration. A mixed-integer programming (MIP) formulation was introduced to identify the best subset of basic terms to lower the degree of expansion and to assist in deriving PCE models. Then, Monte Carlo (MC) simulation was subsequently performed by substituting values of uncertain parameters into closed-form polynomial functions to determine uncertainties of injecting CO_2 underground into a saline aquifer. For each grid at a specific timestep, a PCE model was built to estimate two outcomes: pressure and gas saturation. Uncertain parameters considered included permeability and porosity. Here, 100 numerical simulations were prepared using LHS method to construct many PCEs. This approach was also used to find optimal injection rates with uncertain porosity and permeability.

You *et al.*[234] studied the multi-objective optimisation of a CCUS project located at Andarko Basin, USA. Their work used both weighted sum method[222,234] and Pareto-theory-based optimisation algorithm[235,236] to optimise hydrocarbon production, CO_2 sequestration volume and project economic outcomes simultaneously. The constructed workflow employed ANNs to build robust proxy models and then coupling the proxies with the particle swarm algorithm to carry out the optimisation process. The work emphasised the importance of computationally effective training of ANN proxies and how hyperparameters of trained proxies impact prediction performance. Almasov *et al.*[237] proposed to optimise the design parameters of a single-well CO_2 huff-n-puff process in unconventional oil reservoirs. The optimised objective was to obtain the net present value (NPV) of the process that is estimated using either LS-SVR or GPR. The parameters were optimised using the SQP method. Amar *et al.*[238] introduced a method to optimise the parameters of the CO_2-WAG process to maximise oil production. SVR was used to build the proxy model and then the proxy was used with the GA to find the combinations of parameters that led to the optimal oil production. GA was also utilised as the approach to optimise the hyperparameters of SVR for better proxy performance.

Nwachukwu *et al.*[239] coupled the XGBoost model with a modified version of Mesh Adaptive Direct Search (MADS) to deal with well placement and control optimization in a CO_2-WAG project to obtain maximal NPV. MADS is a pattern search-based method. In the modified MADS, a multidirectional pooling scheme was employed within every iteration to increase the search efficiency. More importantly, the author introduced

a method to reduce the uncertainty existing in the optimised solutions. Since the proxy model will have prediction errors compared with the numerical model, an error model was constructed as a function of control parameters and objective functions (*i.e.*, well placement, water/gas injection rates and NPV) based on the training information. In the optimisation process, if the difference between two candidate optimal solutions was smaller than the estimated proxy errors using the error model, then the original numerical model would be invoked to determine the ''true value'' of the candidate optimal solutions. This method increased the accuracy of the optimisation and lowered the simulator calls. The optimisation results were compared with the results of joint and sequential schemes using MADS with a full reservoir simulator, it showed that the proposed approach could yield a median error of 0.6% and an R^2 of 0.99.

Ampomah *et al.*[186] introduced a method to handle the co-optimization of the cumulative oil production and CO_2 storage within the Farnsworth Unit (FWU). This work combined these two objectives into a single objective function and assigned a unit weight to each one to reduce computational overhead and accelerating optimisation convergence. The combined objective function was used to find the optimal solution incorporating a quadratic response surface that was generated as the proxy model. The proposed method proves computationally efficient in dealing with the co-optimisation problem. Ampomah *et al.*[240] presented an optimisation under uncertainty workflow to ascertain optimum solution in the presence of geological heterogeneity. A neural network optimisation algorithm was utilised to optimise the multi-objective function both with and without geological uncertainty. This work selected vertical permeability anisotropy (K_v/K_h) as the geological uncertain parameter. A developed risk aversion factor was used to quantify and/or represent the confidence levels to assist in decision making. Ampomah *et al.*[241] presented a performance assessment of storage and corresponding oil recovery utilising a Latin hypercube sampling technique to access sensitivity of uncertain parameters towards the pre-defined objective function. A response surface model was constructed using Box–Behnken (BB) deterministic sampling algorithm. A total of 49 simulations were required for training data using this BB design. Forty-nine additional simulations were required to validate the constructed polynomial response surface method (PRSM) model using the BB sampling algorithm. This work elaborated a comprehensive reservoir characterisation framework to quantify heterogeneity uncertainty that led to robust prediction of long-term fate of CO_2 stored within a subject reservoir. Bromhal *et al.*[242] introduced a work to summarise how the National Risk Assessment Partnership (NRAP) handles the long-term quantitative risk assessment for carbon storage. NRAP's method was to divide the carbon storage system into components—reservoir, wells, seals, groundwater, atmosphere. And reduced-order models (ROM) were developed for each component using different approaches, such as look up table (LUT), ANNs and PCEs, Polynomial Regression, RBFs,[188] or Response Surface techniques. The ROMs were mostly used to study concentration and pressure information within the reservoir, especially at the reservoir-seal interface during CO_2 injection and for up to 1000 years post-injection period. These pressures and saturations could then be used as input parameters of wellbore or seal leakage models to predict rates and volumes of leakage of CO_2. Different components could be assembled to simulate the entire system within fractions of seconds. The integrated model could also be used to estimate the probability of failure of a carbon storage system with the help of the MC method.

Nwachukwu *et al.*[243] used XGBoost to teach a proxy model learning the structure of inputs-reservoir responses. They also proposed a method to use physical well locations and well-to-well connectivity as the input variables, which increased the prediction accuracy. The Fast-Marching Method (FMM) introduced by Sethian (1996) was used to calculate the propagation of the pressure front and could be expressed as eqn (2):

$$\sqrt{\alpha(\vec{x})}|\nabla\tau(\vec{x})| = 1 \quad (2)$$

where the $\alpha = k/\varnothing\mu c_t$ is the diffusivity, and τ is the diffusive time of flight in the Fourier domain. The diffusive time of flight can be computed given the location of a well to indicate the peak of pressure front to reach any point in the reservoir. It could be obtained by solving the Eikonal equation and used to represent the connectivity between any two points in the reservoir; a higher τ means lower connectivity. The proposed approach was applied on five different scenarios to demonstrate its feasibility, including (i) a homogeneous waterflooding reservoir model with one injection well, (ii) a waterflooding reservoir with channels and two injection wells, (iii) an ensemble of 20 waterflooding reservoirs with two injection wells, (iv) a CO_2-flooding heterogeneous reservoir with two injection wells, and (v) a CO_2-flooding heterogeneous reservoir with spatially-varying initial fluid saturation and three injectors. It was concluded that the proposed method was able to build a suitable alternative to numerical simulations with reasonable accuracy and this method could be used to deal with problems concerning well-placement optimisation.

4.2.2.3 CO_2-Enhanced coalbed methane. CO_2-Enhanced coalbed methane (CO_2-ECBM) takes the dual benefits of sequestering CO_2 in coal seams and displacing the coalbed methane to be produced. The injection of CO_2 in coal seams will induce significant changes in the physical and chemical properties of coal (such as pore structure, strength, elastic modulus, *etc.*), which in turn affects the CO_2 sequestration performance in coal seams.[244] There are few studies relate CO_2-ECBM with ML techniques, but most of those studies apply ML techniques to predict properties of coal and gas, such as coal strength,[244] CO_2/CH_4 adsorption isotherm,[245,246] crack initiation pressure of coal,[247] coal identification,[248] permeability,[249,250] methane production.[251] Yan *et al.*[244] proposed a hybrid artificial intelligence model integrating back propagation neural network (BPNN), GA and adaptive boosting algorithm (AdaBoost) to predict the unconfined compressive strength of coal according to coal rank, CO_2 interaction time, CO_2 interaction temperature and CO_2 saturation pressure. The adsorption behaviour of CO_2

and methane in coal seams plays a pivotal role in determining the storage amount of the injected greenhouse gas. Feng *et al.*[245] employed seven ML algorithms in the prediction of methane adsorption isotherm on coals. Meng *et al.*[246] used the ANN to predict the excess adsorption amount of supercritical CO_2 on coal from the fundamental physicochemical parameters of coal. The ML model was compared with other seven traditional isotherm models. It was concluded the proposed ML model is not limited to the isothermal conditions and does not require excessive tedious experimental. Yan *et al.*[247] used several ML approaches to estimate the crack initiation pressure (CIP) of supercritical CO_2 fracturing (SCDF) in coal samples. BPNN, extreme learning machine (ELM), and SVM were used to construct the relation from inputs (vertical principal stress, horizontal maximum principal stress, horizontal minimum principal stress, fracturing fluid injection rate, fracturing fluid temperature, tensile strength, elastic modulus, and Poisson's ratio) to the output (*e.g.*, CIP). They pointed out that ground stress, fracturing fluid injection rate, and fracturing fluid temperature would have the highest impacts on the CIP of SCDF. Coal permeability is controlled by various parameters such as confining pressures, temperature, gas pressure, effective stresses, and cleat anisotropy. Sharma *et al.*[249] predicted the CO_2 permeability of India coal at varied injection pressure and effective stress using ANFIS. Yan *et al.*[250] compared different SVM-based approaches in the prediction of the change of coal permeability in the CO_2-ECBM process. The inputs consider CO_2 injection pressure, effective stress, temperature, buried depth and coal rank. The model output is CO_2 permeability.

Injecting CO_2 into shale gas reservoirs is also known as one type of CCUS. When the pressure and temperature is high, CO_2 will have a higher adsorption capacity than methane, especially in the micropore volume fraction, thus enhance gas recovery. Researches regarding CO_2 sequestration and shale gas recovery with ML applications focus on the prediction of kerogen components and types,[252] methane/CO_2 adsorption capacity,[253–256] and process optimisation.[237] The types, molecular components, and structures of shale kerogen directly influence its adsorption and hydrocarbon generation. Kang *et al.*[9] proposed a method to combine ML with nuclear magnetic resonance (NMR) spectra to predict the kerogen components and types in shale. NMR spectrum was used as the inputs since the kerogen molecule's carbon skeleton information was mainly concerned.[256] The 2D spectrum was firstly converted into a 1D matrix where the values representing the NMR spectrum's normalized values, and then was fed into fully connected neural networks (FCNNs). The outputs of the FCNNs were molecular structure labels corresponding to different NMR spectrums. They concluded this method gives excellent performance in the prediction of kerogen skeleton components and types. Meng *et al.*[253] utilised classical approaches and ML approaches in the forecasting of the methane adsorption in shale. Amar *et al.*[254] applied gene expression programming (GEP) and group method of data handling (GMDH) to predict methane adsorption in shale gas formations. The pressure, temperature, total organic carbon, and moisture were considered as input parameters, while gas content (expressed in SCF per ton) was the models' single output. Bemani *et al.*[255] estimated the adsorption capacity of CO_2, CH_4 and CO_2/CH_4 mixture in shale through an ML-based approach. They utilised the LS-SVM to mimic the relationship between four inputs (pressure, temperature, gas composition and TOC) to the gas adsorption capacity. Wang *et al.*[256] utilised different ML algorithms to predict the adsorbed shale gas content using reservoir temperature, TOC, vitrinite reflectance, Langmuir pressure, and Langmuir volume. The methods used include MLR, SVM, RF and ANN. Almasov *et al.*[237] optimized the CO_2 Huff-N-Puff Process in a shale oil reservoir. The NPV was calculated using proxies trained through LS-SVR and GPR. The well control parameters were then optimized to have the optimal NPV.

4.2.2.4 Chemicals, fuels and building materials. CO_2 can be converted into valuable products (chemicals,[257] fuels[258] and building materials[259]) through various physical, chemical or biological pathways.[260] One popular field is CO_2 electrochemical reduction to chemical feedstocks (such as carbon monoxide, formic acid, methanol, methane, ethanol and ethylene) that utilises both CO_2 and hydrogen from renewable energy, to achieve a circular economy.[261] Catalyst development is one of the key steps to realise selective, fast, and efficient reduction processes of CO_2 into valuable products.[262] The ML algorithms showed great advances in efficiently screening the huge number of catalysts for the CO_2 catalytic or electro-catalytic conversion. Ulissi *et al.*[263] proposed to use a neural-network-based surrogate model together with DFT calculations to enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO_2 reduction. Recently, Zhong *et al.*[264] claimed that Cu–Al electrocatalysts can efficiently convert CO_2 to ethylene with the highest faradaic efficiency reported so far through ML and DFT calculations. A ML-augmented chemisorption model has also been proven to be an effective way for CO_2 electroreduction to valuable C_2 species.[265,266] Wu *et al.*[267] found that the computational time and prediction errors could be reduced significantly by employing an extreme GBR. Herein, 80 adsorbate–pair combinations were identified to simultaneously enhance CH_4 and C_2 production on copper after screening 289 combinations. Wan *et al.*[268] also proved that GBR model exhibited the best prediction performance to select the superior electrocatalysts for CO_2 reduction. Moreover, Chen *et al.*[269] developed a ML model based on an extreme gradient boosting regression algorithm and simple features, which can successfully and rapidly predict the Gibbs free energy change of CO adsorption of 1060 atomically dispersed metal–nonmetal co-doped graphene systems, and significantly decrease time and costs. The ML methods show a great potential in accelerating the catalyst development based on the existing experimental results.[270] Li *et al.*[271] evaluated five ML algorithms (SVM, KNN, DT, SGD and ANN) trained by experimental data to classify the characteristics and performance of MOFs for fixing carbon dioxide into cyclic carbonate. The results indicated the six best metal ions (Mn, V, Cu, Ni, Zr and Y) and four best ligands (tactmb, tdcbpp, TCPP, H3L) for new MOFs catalysts for carbon dioxide fixation. In addition, biological fixation is also an

attractive method to convert CO_2 into organic compounds by using organisms such as microalgae. Most of the work are focused on experimental investigations of the CO_2 conversion or utilisation efficiency.[272] Recently, Coşgun *et al.*[273] studied the effect of CO_2 content on the lipid production performance by ML. They indicates that ML is helpful to determine the optimum cultivation conditions and guide for the future scale-up. Thus, the ML approaches should be further applied in the biofixation processes to identify the best CO_2 fixation rate and provide the most beneficial products.

CO_2 can also be utilised to produce the building materials through CO_2 mineralisation. Machine learning is a powerful tool to predict the durability and performance of concrete. Taffese *et al.*[274] applied ANN, DT and ensemble methods to predict the carbonation depth with rationally low error, and the ML models indicated that the CaPrM model can help designers to optimise the concrete mix or structural design as well as to define proactive maintenance plan. Song *et al.*[275] developed a machine-learning-aided platform (ANNs) to enable the rapid, accurate, and high-throughput screening of fly ashes by predicting a structure-based proxy for their reactivity solely on the basis of bulk chemical composition, which has potential to maximise the beneficial utilisation of fly ashes such as CO_2 adsorbents and construction materials.

4.2.3 Perspectives and prospects. ML has been widely applied in CO_2 storage and CO_2-EOR projects. ML was utilised accompanied with numerical simulation to assess the effects of trapping mechanisms on how CO_2 plume spreads and migrates in the underground structure. Several researches focused on CO_2 solubility in oleic and aqueous phases. Various ML algorithms has been employed to investigate relation between CO_2 solubility and factors such as diffusivity, oil/gas–brine IFT, temperature, pressure and brine salinity.

One critical reason for the employment of ML technologies is to construct input–output relations when some critical information is missed or fundamental theory is unclear, which is challenging through traditional approaches. Studies have been performed on how to monitor and detect CO_2 leakage in CCS projects using ML techniques with direct or in-direct monitoring data. The data used include seismic data, downhole monitoring information (such as pressure or TDS), porosity and permeability maps, and injection/production rate, *etc.* Some studies focused on employing ML to predict MMP that is a critical parameter for CO_2-EOR. When coupling CO_2-EOR and CCS, ML-based surrogate models (proxies) have been developed to mimic the original high-fidelity numerical models and to realise part of their functions. This can reduce computational overhead and accelerate exponentially those time-consuming jobs, such as running tens or hundreds of simulations to optimise development schedules or performing uncertainty analysis.

It is important to recognise that ML has been utilised in numerous studies regarding CO_2 storage, utilisation and CO_2-EOR, however, there are still expectations that a more universal workflow will be generated to handle the whole process of a CO_2-EOR-CCS project including data interpretation, storage effect modelling, leakage detection and optimisation jobs, *etc.* Researchers and scientists are also encouraged to study increasing the computational accuracy when building ML-based surrogate models to substitute the original model. Effective use of databases when applying ML warrants further studies.

5. Conclusions

In this work, we have reviewed and discussed the applications of ML in CO_2 capture, transport, storage and utilisation. Firstly, we summarised ML algorithms and suitable platforms that researchers can utilise to accelerate their CCUS research. ML has been extensively applied in both absorbent- and adsorbent-based CO_2 capture processes. For ML in CO_2 absorption, the research is focused on process simulation and optimisation, thermodynamic analysis, and solvent selections and design. As for ML in CO_2 adsorption, the research is focused on applying ML in adsorbent synthesis and characterisation, process modelling and optimisation, and process inversion. It is clear that ML is a powerful tool for screening solvents and adsorbents as well as process modelling and optimisation, which can reduce the development time, capital and operating costs for CO_2 capture. ML is also utilised in oxyfuel combustion for CO_2 capture, in applications such as predictions of combustion characteristics and pollutants emissions and monitoring the combustion process *via* flame images. There are also some studies available that utilise ML models for calcium looping and/or chemical looping combustion for CO_2 capture and this is an area that requires more work. Some researchers have started to apply ML to predict the performance of oxygen carriers and Ca-based sorbents, process control and techno-economic assessment. The experience so far for ML in CO_2 absorption and adsorption, is that it can be adapted to the calcium looping and chemical looping combustion for CO_2 capture. For instance, using QSPR to find the optimal properties of oxygen carriers and Ca-based sorbents for CO_2 separation. ML is also expected to play a vital role in the development of CO_2 utilisation technologies, such as screening catalysts for CO_2 catalytic or electro-catalytic conversion, combined with the DFT calculations, and predicting suitable microalgae types and optimal cultivation conditions for carbon fixation.

ML is also widely applied in CO_2 transportation and storage. It can be incorporated through low-cost sensing techniques to find the hidden relationships in large, complex, and multivariate datasets, to measure the gas–liquid two-phase CO_2 flow with high accuracy and detect leakages during CO_2 transportation. For ML in CO_2 storage, several ML algorithms have been used to investigate the effects of trapping mechanisms on the dispersal and migration of the CO_2 plume, to predict and monitor CO_2 leaking to ensure the safe and long-term storage of injected CO_2 and create the surrogate models for the optimisation of CO_2 CCS-EOR process and uncertainty analysis.

The distinct advantages of applying ML in CCS are that it provides the potential to identify links between data/results that aren't readily identifiable, and it also provides alternative

lower computing cost pathways. Researchers in CCS can apply ML to accelerate the design and development of materials for CO_2 separation and conversion, measure the multiphase CO_2 flow, evaluate the trapping mechanisms for CO_2 storage, and develop the surrogate model for process optimisation and uncertainty analysis. It is important to mention that ML is a data-driven method, which always requires a large quantity of data to develop a generalised and robust model. The quality of training dataset, the selections of input–output features and the type of ML algorithms play a vital role to develop a comprehensive model. As mentioned before, researchers have illustrated suitable methods for feature selection, avoiding the overfitting, and issues with small datasets, when applying ML in CCUS. With the development of ML in CCUS, it is expected that ML will be an efficient and vital tool to accelerate the development of cost-effective CCUS systems to tackle the climate change.

5.1 Overarching perspectives

The authors make the following recommendations to the community for future work and research to increase the take up of CCUS and encourage the development of ML in this field:

(1) Education of ML and CCUS. The education of future generations in ML techniques and CCUS at undergraduate and graduate levels is important and something that is not always part of mainstream curriculums in engineering courses. We therefore recommend ML and CCUS take a greater role in Higher Education practices.

(2) Models should be generalised. Greater emphasis should be placed on transferable learning-focused methods, so that models do not need to be retrained for each material and/or process. Generalised models, which can infer functional information should be explored in CCUS.

(3) Models should offer a combined approach. The development of combined models for materials and process and systems optimisation (performed simultaneously) would prove useful for deployment of CCUS technologies at commercial scale. Most applications of ML so far have been limited to evaluating the technical performance of various processes. Efforts should be made to extend these to incorporate economic, safety and reliability aspects, particularly through techno-economic and life-cycle assessments.

(4) Models need to be tested at scale. More detailed investigations on the effect of process scale (in capture/utilisation) need to be performed. We need to know whether models/designs/optimisation conducted at lab/pilot scales hold at industrial scales, or will models need to be retrained and optimisation redone during scaling up? Can ML models be truly multi-scale (accounting for chemical properties of materials to overall reactor performance) in their CCUS applications? This information will be needed to increase collaboration with industrial partners.

(5) Models need to compensate for lack of data. Further develop hybrid ML methods that find ways to incorporate intuition/domain knowledge to compensate for a lack of data.

(6) Models should go beyond black-boxes. Develop models that are interpretable and explainable, otherwise there is a risk of a lack of trust and acceptability in their take up.

(7) Process control models need developing. Process control is challenging in many CCUS (and other chemical) processes, more work needs to be conducted to understand if ML can be applied to improve process control.

(8) Data and models should be open. We recommend that when ML research is conducted in CCUS, then the training data and ML models should be made publicly accessible in the open domain to enable greater take up and deployment.

(9) Scale up CCUS and use ML where possible. As a final statement, the Paris Agreement and the latest IPCC 6th working group report provide the impetus for both CCUS deployment at scale and harnessing ML to optimise and improve the performance of CCUS technologies. We do not have much time to mitigate the worst effects of climate change, and therefore we must move from CCUS concepts to full scale plants as soon as possible, and ML will be a key enabler of this goal.

Author contributions

YY: conceptualization, visualization, writing – original draft, and writing – review and editing. TNB, SGS, KNP, VP, AR, PN, JOA, ZZ, DS, LW, WZ, YY, WA, JY and MW: writing – original draft, and writing – review and editing. PTC, VM and EJA: conceptualization, project administration, funding acquisition, supervision, and writing – review and editing.

Data availability and access

No new data was generated throughout this review.

Conflicts of interest

There are no conflicts to declare.

Nomenclature

$K_G a$	Mass-transfer coefficient
ω	Acentric factor
Na-Phe	Sodium salt of L-phenylalanine
3DMA1P	3-Dimethylamino-1-propanol
R^2	Coefficient of determination
C_P	Heat capacity
α	Diffusivity
τ	Diffusive time of flight in the Fourier domain

Abbreviations

AAD	Average absolute deviation
AARD%	Average absolute relative deviation in percent
ANFIS	Adaptive network-based fuzzy inference system
ABC	Artificial bee colony
AE	Acoustic emission

AI	Artificial intelligence
ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average
BPNN	Back-propagation neural network
BR	Bayesian regularization
BECCS	Biomass energy with carbon capture and storage
BB	Box–Behnken
CCS	Carbon capture and storage
CCUS	Carbon capture, utilisation and storage
CLC	Chemical-looping combustion
CFB	Circulating fluidised bed
CART	Classification and regression tree
CAMD	Computer-aided molecular design
CSA	Concentration swing adsorption
CNN	Convolutional neural networks
COFs	Covalent organic frameworks
CSS	Cyclic-steady state
DFT	Density functional theory
DTs	Decision trees
DBN	Deep belief network
DNN	Deep neural network
DQN	Deep Q network
DQL	Deep Q-learning
DBSCAN	Density-based spatial clustering of applications with noise
DOE	Design of experiments
DAC	Direct air capture
DFB	Dual-fluidised bed
ESA	Electric swing adsorption
EOR	Enhanced oil recovery
EOS	Equation of state
EU-ETS	European Union Emission Trading Scheme
ELM	Extreme learning machine
FMM	Fast-marching method
FWU	Farnsworth unit
GRU	Gate recurrent unit
GPR	Gaussian process regression
GRNN	General regression neural network
GA	Genetic algorithm
GA-ANN	Genetic algorithm-artificial neural network
GP	Genetic programming
GCS	Geological carbon sequestration
GBR	Gradient boosted regression
GCMC	Grand Canonical Monte Carlo
SRM_G	Grid-based surrogate reservoir model
GC	Group contribution
GMDH	Group method of data handling
GRU	Gate recurrent unit
GVF	Gas volume fraction
HPT	Harmonic pulse testing
HDMR	High dimensional model representation
IPC	Intelligent predictive controller
IFT	Interfacial tension
ILs	Ionic liquids
KSVM	Kernel support vector machine
LHS	Latin hypercube sampling
LS-SVM	Least square support vector machine
LSTM	Long short-term memory
LMA	Levenberg–Marquardt algorithm
LUT	Look up table
ML	Machine learning
MAPLE	Machine-assisted adsorption process learner and emulator
MDP	Markov decision process
MAE	Mean absolute error
MERQ	Material, energy, rate and equilibrium
MADS	Mesh adaptive direct search
MESH	Mass, equilibrium summation and enthalpy
MOFs	Metal–organic frameworks
MSA	Microwave swing adsorption
MMP	Minimum miscibility pressure
MIP	Mixed-integer programming
MD	Molecular dynamics simulations
MM	Molecular mechanically
MEA	Monoethanolamine
MC	Monte Carlo
MLP	Multi-layer perceptron
MLP-ANN	Multi-layer perceptron artificial neural network
MLP-LMA	Multi-layer perceptron Levenberg–Marquardt algorithm
MARS	Multivariate adaptive regression splines
MLR	Multivariate linear regression
NRAP	National risk assessment partnership
NDCs	Nationally determined contributions
NET	Negative emissions technologies
MDEA	*N*-Methyl diethanolamine
NLP	Non-linear programming
PDE	Partial differential equations
PLS	Partial least-squares
PLSR	Partial least squares regression
PMI	Partial mutual information
PSO	Particle swarm optimization
PI	Permutation importance
PZ	Piperazine
PCE	Polynomial chaos expansion
PRSM	Polynomial response surface method
POCs	Porous organic cages
PSA	Pressure swing adsorption
PCA	Principal component analysis
PCA–RWN	Principal component analysis and random weight network
PCR	Principal component regression
QSPR/QSAR	Quantitative-structure property/activity relationship
QM	Quantum-mechanically
RBF	Radial basis function
RBFNN	Radial basis function neural network
RF	Random forest
RNN	Recurrent neural network
ROM	Reduced-order models
RL	Reinforcement learning
ROZ	Residual oil zones

RSM	Response surface methodology
RMSE	Root mean square error
RPB	Rotating packed bed
SCG	Scaled conjugate gradient
SQP	Sequential quadratic programming algorithm
SMR	Steam methane reforming
SE-SMR	Sorption enhanced steam methane reforming
SVM	Support vector machine
SVMr	SVM with a radial basis kernel
SVR	Support vector regression
TLBO-RBF	Teaching–learning-based optimization hybrid model
TRL	Technology readiness level
TSA	Temperature swing adsorption
TVSA	Temperature-vacuum swing adsorption
TGA	Thermogravimetric analysis
TDS	Total dissolved solids
VSA	Vacuum swing adsorption
VLE	Vapour–liquid equilibrium
WAG	Water alternating gas
SRM_W	Well-based surrogate reservoir model
ZIFs	Zeolitic imidazolate frameworks

Acknowledgements

PN, PTC, and VM acknowledge the financial support from the UK Engineering and Physical Sciences Research Council Doctoral Training Partnership (EPSRC DTP) grant no. EP/R513027/1. Yongliang Yan would like to acknowledge the financial support from the Cranfield University Energy and Power research bursary. JOA is grateful to the Petroleum Technology Development Fund (PTDF), Nigeria, for doctoral study scholarship, award number: PTDF/ED/OSS/PHD/JOA/077/19 and the University of Benin, Benin City, Nigeria. WZ and Yong Yan acknowledge the financial support from the National Natural Science Foundation of China (No. 61973113 and No. 62073135). WA and JY would like to acknowledge grant support provided by the U.S. Department of Energy's (DOE) National Energy Technology Laboratory (NETL) through the Southwest Regional Partnership on Carbon Sequestration (SWP) under Award No. DE-FC26-05NT42591. SGS, KNP, VP and AR acknowledge funding from Canada First Research Excellence Fund through University of Alberta Future Energy systems, MW would like to thank the financial support from UK EPSRC (EP/M001458/2 and EP/N024540/1).

References

1 Earth's CO_2 Home Page, https://www.co2.earth/ (accessed 28 April 2021).
2 United Nations, The Greatest Threat To Global Security: Climate Change Is Not Merely An Environmental Problem | United Nations, https://www.un.org/en/chronicle/article/greatest-threat-global-security-climate-change-not-merely-environmental-problem (accessed 21 July 2021).
3 HOME – UN Climate Change Conference (COP26) at the SEC – Glasgow 2021, https://ukcop26.org/ (accessed 28 April 2021).
4 Build Back Better: our plan for growth (HTML) – GOV.UK, https://www.gov.uk/government/publications/build-back-better-our-plan-for-growth/build-back-better-our-plan-for-growth-html (accessed 28 April 2021).
5 Executive Order on Protecting Public Health and the Environment and Restoring Science to Tackle the Climate Crisis | The White House, https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-protecting-public-health-and-environment-and-restoring-science-to-tackle-climate-crisis/ (accessed 28 April 2021).
6 Recovery plan for Europe | European Commission, https://ec.europa.eu/info/strategy/recovery-plan-europe_en (accessed 28 April 2021).
7 Executive Order on Tackling the Climate Crisis at Home and Abroad | The White House, https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/executive-order-on-tackling-the-climate-crisis-at-home-and-abroad/ (accessed 28 April 2021).
8 Canada unveils federal budget to end 'Covid recession' | Financial Times, https://www.ft.com/content/2f5befa4-0c6c-4dd6-aea1-657ca23685fc (accessed 28 April 2021).
9 Facilities – Global CCS Institute, https://co2re.co/FacilityData (accessed 28 April 2021).
10 S. Pradhan, W. M. Shobe, J. Fuhrman, H. McJeon, M. Binsted, S. C. Doney and A. F. Clarens, *Front. Clim.*, 2021, **3**, 660787.
11 Net-Zero Emissions by 2050 – Canada.ca, https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/net-zero-emissions-2050.html (accessed 16 July 2021).
12 China's net-zero ambitions: the next Five-Year Plan will be critical for an accelerated energy transition – Analysis – IEA, https://www.iea.org/commentaries/china-s-net-zero-ambitions-the-next-five-year-plan-will-be-critical-for-an-accelerated-energy-transition (accessed 16 July 2021).
13 C. Pozo, Á. Galán-Martín, D. M. Reiner, N. Mac Dowell and G. Guillén-Gosálbez, *Nat. Clim. Change*, 2020, **10**, 640–646.
14 J. Deutch, *Joule*, 2020, **4**, 2237–2240.
15 C. T. M. Clack, S. A. Qvist, J. Apt, M. Bazilian, A. R. Brandt, K. Caldeira, S. J. Davis, V. Diakov, M. A. Handschy, P. D. H. Hines, P. Jaramillo, D. M. Kammen, J. C. S. Long, M. G. Morgan, A. Reed, V. Sivaram, J. Sweeney, G. R. Tynan, D. G. Victor, J. P. Weyant and J. F. Whitacre, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 6722–6727.
16 M. Bui, C. S. Adjiman, A. Bardow, E. J. Anthony, A. Boston, S. Brown, P. S. Fennell, S. Fuss, A. Galindo, L. A. Hackett, J. P. Hallett, H. J. Herzog, G. Jackson, J. Kemper, S. Krevor, G. C. Maitland, M. Matuszewski, I. S. Metcalfe, C. Petit, G. Puxty, J. Reimer, D. M. Reiner, E. S. Rubin, S. A. Scott, N. Shah, B. Smit, J. P. M. Trusler, P. Webley, J. Wilcox and N. Mac Dowell, *Energy Environ. Sci.*, 2018, **11**, 1062–1176.
17 Y. Yan, T. Mattisson, P. Moldenhauer, E. J. Anthony and P. T. Clough, *Chem. Eng. J.*, 2020, **387**, 124072.

18 Z. Bai, F. Li, J. Zhang, E. Oko, M. Wang, Z. Xiong and D. Huang, *Comput. Aided Chem. Eng.*, 2016, **38**, 2007–2012.
19 F. Li, J. Zhang, E. Oko and M. Wang, *Int. J. Coal Sci. Technol.*, 2017, **4**, 33–40.
20 F. Li, J. Zhang, C. Shang, D. Huang, E. Oko and M. Wang, *Appl. Therm. Eng.*, 2018, **130**, 997–1003.
21 Z. Li, Z. Ding and M. Wang, *Engineering*, 2017, **3**, 257–265.
22 X. Wu, M. Wang, J. Shen, Y. Li, A. Lawal and K. Y. Lee, *Appl. Energy*, 2019, **238**, 495–515.
23 X. Wu, J. Shen, M. Wang and K. Y. Lee, *Energy*, 2020, **196**, 117070.
24 A. H. Farmahini, S. Krishnamurthy, D. Friedrich, S. Brandani and L. Sarkisov, *Ind. Eng. Chem. Res.*, 2018, **57**(45), 15491–15511.
25 T. D. Burns, K. N. Pai, S. G. Subraveti, S. P. Collins, M. Krykunov, A. Rajendran and T. K. Woo, *Environ. Sci. Technol.*, 2020, **54**, 4536–4544.
26 K. N. Pai, V. Prasad and A. Rajendran, *Ind. Eng. Chem. Res.*, 2020, **59**, 16730–16740.
27 D. Shao, Y. Yan, W. Zhang, S. Sun, C. Sun and L. Xu, *Int. J. Greenh. Gas Control*, 2020, **94**, 102950.
28 S. Sinha, R. P. de Lima, Y. Lin, A. Y. Sun, N. Symon, R. Pawar and G. Guthrie, Proceedings – SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers (SPE), 2020, vol. 2020 – October.
29 M. I. Jordan and T. M. Mitchell, *Science*, 2015, **349**, 255–260.
30 D. M. J. Garbade, Clearing the Confusion: AI vs Machine Learning vs Deep Learning Differences | by Dr Michael J. Garbade | Towards Data Science, https://towardsdatascience.com/clearing-the-confusion-ai-vs-machine-learning-vs-deep-learning-differences-fce69b21d5eb (accessed 7 April 2021).
31 Sunil Kumar, Advantages and Disadvantages of Artificial Intelligence | by sunil kumar | Towards Data Science, https://towardsdatascience.com/advantages-and-disadvantages-of-artificial-intelligence-182a5ef6588c (accessed 7 April 2021).
32 M. Stewart, The Limitations of Machine Learning | by Matthew Stewart, PhD Researcher | Towards Data Science, https://towardsdatascience.com/the-limitations-of-machine-learning-a00e0c3040c6 (accessed 7 April 2021).
33 J. T. Raj, A beginner's guide to dimensionality reduction in Machine Learning | by Judy T Raj | Towards Data Science, https://towardsdatascience.com/dimensionality-reduction-for-machine-learning-80a46c2ebb7e.
34 L. J. P. Van Der Maaten, E. O. Postma and H. J. Van Den Herik, *J. Mach. Learn. Res.*, 2009, **10**, 1–41.
35 S. Khalid, T. Khalil and S. Nasreen, Proc. 2014 Sci. Inf. Conf. SAI 2014, 2014, pp. 372–378.
36 R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti and D. Pedreschi, *ACM Comput. Surv.*, 2018, **51**(5), 93, DOI: 10.1145/3236009.
37 H. Patrick and G. Navdeep, *An Introduction to Machine Learning Interpretability*, 2019, vol. 1.
38 S. Das, N. Agarwal, D. Venugopal, F. T. Sheldon and S. Shiva, 2020 IEEE Symp. Ser. Comput. Intell. SSCI 2020, 2020, pp. 670–677.
39 A. Singh, N. Thakur and A. Sharma, *A review of supervised machine learning algorithms*, IEEE, 2016, pp. 1310–1315.
40 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.
41 S. Bhatt, Reinforcement Learning 101. Learn the essentials of Reinforcement, https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292 (accessed 7 April 2021).
42 P. Pareek, Machine learning Algorithms and where they are used?, https://medium.com/@priyapareek0205/machine-learning-algorithms-and-where-they-are-used-c74de1441e1 (accessed 7 April 2021).
43 Y. Yan, T. N. Borhani and P. T. Clough, *RSC Theoretical and Computational Chemistry Series*, Royal Society of Chemistry, 2020, vol. 2020-January, pp. 340–371.
44 T. N. G. Borhani, A. Azarpour, V. Akbari, S. R. Wan Alwi and Z. A. Manan, *Int. J. Greenh. Gas Control*, 2015, **41**, 142–162.
45 T. N. Borhani, E. Oko and M. Wang, *J. Clean. Prod.*, 2018, **204**, 1124–1142.
46 N. Sipöcz, F. A. Tobiesen and M. Assadi, *Appl. Energy*, 2011, **88**, 2368–2376.
47 S. Babamohammadi, A. Shamiri, T. Nejad Ghaffar Borhani, M. S. Shafeeyan, M. K. Aroua and R. Yusoff, *J. Mol. Liq.*, 2018, **249**, 40–52, DOI: 10.1016/j.molliq.2017.10.151.
48 A. Nuchitprasittichai and S. Cremaschi, *Ind. Eng. Chem. Res.*, 2013, **52**, 10236–10243.
49 F. Li, J. Zhang, E. Oko and M. Wang, *Fuel*, 2015, **151**, 156–163.
50 J. Zhan, B. Wang, L. Zhang, B.-C. Sun, J. Fu, G. Chu and H. Zou, *Ind. Eng. Chem. Res.*, 2020, **59**, 8295–8303.
51 A. Shalaby, A. Elkamel, P. L. Douglas, Q. Zhu and Q. P. Zheng, *Energy*, 2021, **215**, 119113.
52 T. N. Borhani, S. A. Nabavi, D. P. Hanak and V. Manovic, *Rev. Chem. Eng.*, 2021, **37**(8), 931–957.
53 A. Baghban, A. Bahadori, A. H. Mohammadi and A. Behbahaninia, *Int. J. Greenh. Gas Control*, 2017, **57**, 143–161.
54 A. Dey, *Int. J. Comput. Sci. Inf. Technol.*, 2016, **7**, 1174–1179.
55 M. M. Ghiasi and A. H. Mohammadi, *J. Mol. Liq.*, 2017, **242**, 594–605.
56 S. Garg, A. M. Shariff, M. S. Shaikh, B. Lal, H. Suleman and N. Faiqa, *J. CO2 Util.*, 2017, **19**, 146–156.
57 C. Li, H. Liu, M. Xiao, X. Luo, H. Gao and Z. Liang, *Int. J. Greenh. Gas Control*, 2017, **63**, 77–85.
58 D. M. Austgen, G. T. Rochelle and C. C. Chen, *Ind. Eng. Chem. Res.*, 1991, **30**, 543–555.
59 W. Hu and A. Chakma, *Chem. Eng. Commun.*, 1990, **94**, 53–61.
60 H. Liu, M. Xiao, Z. Liang, W. Rongwong, J. Li and P. Tontiwachwuthikul, *Ind. Eng. Chem. Res.*, 2015, **54**, 12525–12533.
61 S. Babamohammadi, A. Shamiri, T. Nejad Ghaffar Borhani, M. S. Shafeeyan, M. K. Aroua and R. Yusoff, *J. Mol. Liq.*, 2018, **249**, 40–52.
62 H. Yarveicy, H. Saghafi, M. M. Ghiasi and A. H. Mohammadi, *Environ. Prog. Sustainable Energy*, 2019, **38**, S441–S448.

63 E. Soroush, M. Mesbah, N. Hajilary and M. Rezakazemi, *J. Environ. Chem. Eng.*, 2019, **7**, 102925.
64 T. E. Daubert, R. P. D. R. L. Rowley, W. V. Wilding, J. L. Oscarson, Y. Yang and N. A. Zundel, *Des. Inst. Phys. Prop.*, 2003, https://scholar.google.com/citations?view_op=view_citation&hl=en&user=MgOnMwYAAAAJ&citation_for_view=MgOnMwYAAAAJ:k_IJM867U9cC.
65 P. J. Linstrom and W. G. Mallard, *J. Chem. Eng. Data*, 2001, **46**, 1059–1063.
66 U. Westhaus, T. Droge and R. Sass, *Fluid Phase Equilibria*, Elsevier, 1999, vol. 158–160, pp. 429–435.
67 B. E. Poling, J. M. Prausnitz, J. P. O'connell, N. York, C. San, F. Lisbon, L. Madrid, M. City, M. N. Delhi and S. Juan, *The properties of gases and liquids*, McGraw-Hill Education, 5th edn, 2001.
68 V. Papaioannou, C. S. Adjiman, G. Jackson and A. Galindo, *Process Systems Engineering*, John Wiley & Sons, Ltd, 2011, pp. 135–172.
69 T. N. Borhani, S. García-Muñoz, C. Vanesa Luciani, A. Galindo and C. S. Adjiman, *Phys. Chem. Chem. Phys.*, 2019, **21**, 13706–13720.
70 K. Golzar, S. Amjad-Iranagh and H. Modarress, *Meas. J. Int. Meas. Confed.*, 2013, **46**, 4206–4225.
71 V. Venkatraman and B. K. Alsberg, *J. CO2 Util.*, 2017, **21**, 162–168.
72 M. A. Kuenemann and D. Fourches, *Mol. Inform.*, 2017, **36**, 1600143.
73 Z. Zhang, H. Li, H. Chang, Z. Pan and X. Luo, *J. CO2 Util.*, 2018, **26**, 152–159.
74 M. Afkhamipour, M. Mofarahi, T. N. G. Borhani and M. Zanganeh, *Heat Mass Transfer*, 2018, **54**, 855–866.
75 L. Cao, P. Zhu, Y. Zhao and J. Zhao, *J. Hazard. Mater.*, 2018, **352**, 17–26.
76 T. N. Borhani, E. Oko and M. Wang, *J. Ind. Eng. Chem.*, 2019, **75**, 285–295.
77 S. A. Mazari, A. R. Siyal, N. H. Solangi, S. Ahmed, G. Griffin, R. Abro, N. M. Mubarak, M. Ahmed and N. Sabzoi, *J. Mol. Liq.*, 2020, 114785.
78 T. Wu, W. L. Li, M. Y. Chen, Y. M. Zhou and Q. Y. Zhang, *Chem. Pap.*, 2021, **75**, 1619–1628.
79 T. N. Borhani and M. Wang, *Renewable Sustainable Energy Rev.*, 2019, 114.
80 M. Fujinami, H. Maekawara, R. Isshiki, J. Seino, J. Yamaguchi and H. Nakai, *Bull. Chem. Soc. Jpn.*, 2020, **93**, 841–845.
81 T. Zhou, K. McBride, S. Linke, Z. Song and K. Sundmacher, *Curr. Opin. Chem. Eng.*, 2020, **27**, 35–44.
82 Y. Liu, H. Yu, Y. Sun, S. Zeng, X. Zhang, Y. Nie, S. Zhang and X. Ji, *Front. Chem.*, 2020, **8**, 82.
83 M. Taheri, R. Zhu, G. Yu and Z. Lei, *Chem. Eng. Sci.*, 2021, **230**, 116199.
84 V. Venkatraman, S. Evjen, K. C. Lethesh, J. J. Raj, H. K. Knuutila and A. Fiksdahl, *Sustainable Energy Fuels*, 2019, **3**, 2798–2808.
85 K. Wang, H. Xu, C. Yang and T. Qiu, *Green Energy Environ.*, 2021, **6**(3), 432–443.
86 Principles of Adsorption and Adsorption Processes - Douglas M. Ruthven – Google Books, https://books.google.co.uk/books?hl=en&lr=&id=u7wq21njR3UC&oi=fnd&pg=PR17&dq=1.%09D.+M.+Ruthven,+Principles+of+adsorption+and+adsorption+processes,+John+Wiley,+New+York,+1984.&ots=wcTuRuzjxX&sig=qoElCOm1JqWpmOakwxFzmJzy2Qo#v=onepage&q=1.%2509D.M.Ruthven%252(accessed 7 April 2021).
87 D. Ruthven, F. P.-N. Y. V. Publishers and undefined 1994, *sutlib2.sut.ac.th*.
88 J. R. Long and O. M. Yaghi, *Chem. Soc. Rev.*, 2009, **38**, 1213–1214.
89 J. R. Li, J. Sculley and H. C. Zhou, *Chem. Rev.*, 2012, **112**, 869–932.
90 O. M. Yaghi and H. Li, *J. Am. Chem. Soc.*, 1995, **117**, 10401–10402.
91 K. Geng, T. He, R. Liu, S. Dalapati, K. T. Tan, Z. Li, S. Tao, Y. Gong, Q. Jiang and D. Jiang, *Chem. Rev.*, 2020, **120**, 8814–8933.
92 B. Chen, Z. Yang, Y. Zhu and Y. Xia, *J. Mater. Chem. A*, 2014, **2**, 16811–16831.
93 T. Hasell and A. I. Cooper, *Nat. Rev. Mater.*, 2016, **1**, 1–14.
94 L. C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk and B. Smit, *Nat. Mater.*, 2012, **11**, 633–641.
95 T. M. McDonald, J. A. Mason, X. Kong, E. D. Bloch, D. Gygi, A. Dani, V. Crocellà, F. Giordanino, S. O. Odoh, W. S. Drisdell, B. Vlaisavljevich, A. L. Dzubak, R. Poloni, S. K. Schnell, N. Planas, K. Lee, T. Pascal, L. F. Wan, D. Prendergast, J. B. Neaton, B. Smit, J. B. Kortright, L. Gagliardi, S. Bordiga, J. A. Reimer and J. R. Long, *Nature*, 2015, **519**, 303–308.
96 R. L. Siegelman, T. M. McDonald, M. I. Gonzalez, J. D. Martell, P. J. Milner, J. A. Mason, A. H. Berger, A. S. Bhown and J. R. Long, *J. Am. Chem. Soc.*, 2017, **139**, 10526–10538.
97 A. S. Rosen, M. R. Mian, T. Islamoglu, H. Chen, O. K. Farha, J. M. Notestein and R. Q. Snurr, *J. Am. Chem. Soc.*, 2020, **142**, 4317–4328.
98 S. M. Moosavi, K. M. Jablonka and B. Smit, *J. Am. Chem. Soc.*, 2020, **142**, 20273–20287.
99 K. M. Jablonka, D. Ongari, S. M. Moosavi and B. Smit, *Chem. Rev.*, 2020, **120**, 8066–8129.
100 C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, *Nat. Chem.*, 2012, **4**, 83–89.
101 P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward and D. Fairen-Jimenez, *Chem. Mater.*, 2017, **29**, 2618–2625.
102 P. G. Boyd and T. K. Woo, *CrystEngComm*, 2016, **18**, 3777–3792.
103 Y. J. Colón, D. A. Gómez-Gualdrón and R. Q. Snurr, *Cryst. Growth Des.*, 2017, **17**, 5801–5810.
104 J. C. Cole, C. R. Groom, M. G. Read, I. Giangreco, P. McCabe, A. M. Reilly and G. P. Shields, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 530–541.
105 Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling,

J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl and R. Q. Snurr, *J. Chem. Eng. Data*, 2019, **64**, 5985–5998.
106 R. Pophale, P. A. Cheeseman and M. W. Deem, *Phys. Chem. Chem. Phys.*, 2011, **13**, 12407–12412.
107 N. Stock and S. Biswas, *Chem. Rev.*, 2012, **112**, 933–969.
108 D. J. Tranchemontagne, J. R. Hunt and O. M. Yaghi, *Tetrahedron*, 2008, **64**, 8553–8557.
109 V. Subramanian Balashankar and A. Rajendran, *ACS Sustainable Chem. Eng.*, 2019, **7**, 17747–17755.
110 K. T. Leperi, Y. G. Chung, F. You and R. Q. Snurr, *ACS Sustainable Chem. Eng.*, 2019, **7**, 11529–11539.
111 M. Khurana and S. Farooq, *AIChE J.*, 2019, **65**, 184–195.
112 A. H. Farmahini, S. Krishnamurthy, D. Friedrich, S. Brandani and L. Sarkisov, *Ind. Eng. Chem. Res.*, 2018, **57**, 15491–15511.
113 P. Z. Moghadam, S. M. J. Rogge, A. Li, C. M. Chow, J. Wieme, N. Moharrami, M. Aragones-Anglada, G. Conduit, D. A. Gomez-Gualdron, V. Van Speybroeck and D. Fairen-Jimenez, *Matter*, 2019, **1**, 219–234.
114 A. Ö. Yazaydin, R. Q. Snurr, T. H. Park, K. Koh, J. Liu, M. D. LeVan, A. I. Benin, P. Jakubczak, M. Lanuza, D. B. Galloway, J. J. Low and R. R. Willis, *J. Am. Chem. Soc.*, 2009, **131**, 18198–18199.
115 S. P. Collins, T. D. Daff, S. S. Piotrkowski and T. K. Woo, *Sci. Adv.*, 2016, **2**, e1600954.
116 P. G. Boyd, A. Chidambaram, E. García-Díez, C. P. Ireland, T. D. Daff, R. Bounds, A. Gładysiak, P. Schouwink, S. M. Moosavi, M. M. Maroto-Valer, J. A. Reimer, J. A. R. Navarro, T. K. Woo, S. Garcia, K. C. Stylianou and B. Smit, *Nature*, 2019, **576**, 253–256.
117 B. Smit and T. L. M. Maesen, *Chem. Rev.*, 2008, **108**, 4125–4184.
118 Q. Yang, D. Liu, C. Zhong and J. R. Li, *Chem. Rev.*, 2013, **113**, 8261–8323.
119 Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl and R. Q. Snurr, *Chem. Mater.*, 2014, **26**, 6185–6192.
120 M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji and T. K. Woo, *J. Phys. Chem. Lett.*, 2014, **5**, 3056–3060.
121 Y. Lee, S. D. Barthel, P. Dłotko, S. M. Moosavi, K. Hess and B. Smit, *Nat. Commun.*, 2017, **8**, 1–8.
122 H. Dureckova, M. Krykunov, M. Z. Aghaji and T. K. Woo, *J. Phys. Chem. C*, 2019, **123**, 4133–4139.
123 J. Burner, L. Schwiedrzik, M. Krykunov, J. Luo, P. G. Boyd and T. K. Woo, *J. Phys. Chem. C*, 2020, **124**, 27996–28005.
124 M. Pardakhti, E. Moharreri, D. Wanik, S. L. Suib and R. Srivastava, *ACS Comb. Sci.*, 2017, **19**, 640–645.
125 B. J. Bucior, N. S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O. K. Farha, N. Bagheri and R. Q. Snurr, *Mol. Syst. Des. Eng.*, 2019, **4**, 162–174.
126 R. Ma, Y. J. Colón and T. Luo, *ACS Appl. Mater. Interfaces*, 2020, **12**, 34041–34048.
127 R. Anderson, A. Biong and D. A. Gómez-Gualdrón, *J. Chem. Theory Comput.*, 2020, **16**, 1271–1283.
128 J. D. Evans and F. X. Coudert, *Chem. Mater.*, 2017, **29**, 7833–7839.
129 K. N. Pai, V. Prasad and A. Rajendran, *Sep. Purif. Technol.*, 2020, **241**, 116651.
130 S. G. Subraveti, Z. Li, V. Prasad and A. Rajendran, *Ind. Eng. Chem. Res.*, 2019, **58**, 20412–20422.
131 J. Xiao, C. Li, L. Fang, P. Böwer, M. Wark, P. Bénard and R. Chahine, *Int. J. Energy Res.*, 2020, **44**, 4475–4492.
132 N. D. Vo, D. H. Oh, J. H. Kang, M. Oh and C. H. Lee, *Appl. Energy*, 2020, **273**, 115263.
133 G. Hüllen, J. Zhai, S. H. Kim, A. Sinha, M. J. Realff and F. Boukouvala, *Comput. Chem. Eng.*, 2020, **136**, 106519.
134 K. T. Leperi, D. Yancy-Caballero, R. Q. Snurr and F. You, *Ind. Eng. Chem. Res.*, 2019, **58**, 18241–18252.
135 L. M. C. Oliveira, H. Koivisto, I. G. I. Iwakiri, J. M. Loureiro, A. M. Ribeiro and I. B. R. Nogueira, *Chem. Eng. Sci.*, 2020, **224**, 115801.
136 A. K. Rajagopalan, A. M. Avila and A. Rajendran, *Int. J. Greenh. Gas Control*, 2016, **46**, 76–85.
137 K. Nagesh Pai, V. Prasad and A. Rajendran, *ACS Sustainable Chem. Eng.*, 2021, **9**(10), 3838–3849.
138 Z. Yao, B. Sánchez-Lengeling, N. S. Bobbitt, B. J. Bucior, S. G. H. Kumar, S. P. Collins, T. Burns, T. K. Woo, O. K. Farha, R. Q. Snurr and A. Aspuru-Guzik, *Nat. Mach. Intell.*, 2021, **3**, 76–86.
139 Q. Zhu, J. M. Jones, A. Williams and K. M. Thomas, *Fuel*, 1999, **78**, 1755–1762.
140 J. Chen, C. Xie, J. Liu, Y. He, W. Xie, X. Zhang, K. Chang, J. Kuo, J. Sun, L. Zheng, S. Sun, M. Buyukada and F. Evrendilek, *Bioresour. Technol.*, 2018, **250**, 230–238.
141 C. Xie, J. Liu, X. Zhang, W. Xie, J. Sun, K. Chang, J. Kuo, W. Xie, C. Liu, S. Sun, M. Buyukada and F. Evrendilek, *Appl. Energy*, 2018, **212**, 786–795.
142 B. Govindan, S. Chandra Babu Jakka, T. K. Radhakrishnan, A. K. Tiwari, T. M. Sudhakar, P. Shanmugavelu, A. K. Kalburgi, A. Sanyal and S. Sarkar, *Energy Fuels*, 2018, **32**, 3995–4007.
143 H. Qiao and S. Zeng, *Pet. Sci. Technol.*, 2019, **37**, 215–219.
144 P. Debiagi, H. Nicolai, W. Han, J. Janicka and C. Hasse, *Fuel*, 2020, 274.
145 J. Krzywanski, T. Czakiert, A. Blaszczuk, R. Rajczyk, W. Muskala and W. Nowak, *Fuel Process. Technol.*, 2015, **137**, 66–74.
146 X. Bai, G. Lu, M. M. Hossain, J. Szuhánszki, S. S. Daood, W. Nimmo, Y. Yan and M. Pourkashanian, *Fuel*, 2017, **202**, 656–664.
147 X. Bai, G. Lu, M. M. Hossain, Y. Yan and S. Liu, *Combust. Sci. Technol.*, 2017, **189**, 776–792.
148 Y. Liu, Y. Fan and J. Chen, *Energy Fuels*, 2017, **31**, 8776–8783.
149 Z. Han, M. M. Hossain, Y. Wang, J. Li and C. Xu, *Appl. Energy*, 2020, 259.
150 N. R. Singstock, C. J. Bartel, A. M. Holder and C. B. Musgrave, *Adv. Energy Mater.*, 2020, **10**, 1–11.
151 Z. T. Wilson and N. V. Sahinidis, *Comput. Chem. Eng.*, 2019, **127**, 88–98.
152 J. Pan, Y. Pottimurthy, D. Wang, S. Hwang, S. Patil and L. S. Fan, *Powder Technol.*, 2020, **367**, 266–276.

153 H. Chen, J. Yan, R. Wei, J. Gao, J. Lian and X. Huang, Asia-Pacific Power and Energy Engineering Conference, APPEEC, 2011.
154 P. Nkulikiyinka, Y. Yan, F. Güleç, V. Manovic and P. T. Clough, *Energy AI*, 2020, **2**, 100037.
155 J. Krzywanski, T. Czakiert, T. Shimizu, I. Majchrzak-Kuceba, Y. Shimazaki, A. Zylka, K. Grabowska and M. Sosnowski, *Energy Fuels*, 2018, **32**, 6355–6362.
156 D. P. Hanak and V. Manovic, *Appl. Energy*, 2017, **208**, 691–702.
157 J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti and M. A. L. Marques, *Chem. Mater.*, 2017, **29**, 5090–5103.
158 J. W. Barnett, C. R. Bilchak, Y. Wang, B. C. Benicewicz, L. A. Murdock, T. Bereau and S. K. Kumar, *Sci. Adv.*, 2020, **6**, eaaz4301.
159 S. Ullah, M. A. Assiri, A. G. Al-Sehemi, M. A. Bustam, H. Abdul Mannan, F. A. Abdulkareem, A. Irfan and S. Saqib, *Greenhouse Gases: Sci. Technol.*, 2019, **9**, 1010–1026.
160 A. L. Ahmad, J. K. Adewole, C. P. Leo, S. Ismail, A. S. Sultan and S. O. Olatunji, *J. Membr. Sci.*, 2015, **480**, 39–46.
161 V. E. Onyebuchi, A. Kolios, D. P. Hanak, C. Biliyok and V. Manovic, *Renewable Sustainable Energy Rev.*, 2018, **81**, 2563–2583.
162 NCCS, NCCS Annual report 2019.
163 H. Li, B. Dong, Z. Yu, J. Yan and K. Zhu, *Appl. Energy*, 2019, **255**, 113789.
164 I. S. Cole, P. Corrigan, S. Sim and N. Birbilis, *Int. J. Greenh. Gas Control*, 2011, **5**, 749–756.
165 G. J. Collie, M. Nazeri, A. Jahanbakhsh, C.-W. Lin and M. M. Maroto-Valer, *Greenh. Gases Sci. Technol.*, 2017, **7**, 10–28.
166 E. Nazemi, S. A. H. Feghhi, G. H. Roshani, R. Gholipour Peyvandi and S. Setayeshi, *Nucl. Eng. Technol.*, 2016, **48**, 64–71.
167 K. T. O'Neill, L. Brancato, P. L. Stanwix, E. O. Fridjonsson and M. L. Johns, *Chem. Eng. Sci.*, 2019, **202**, 222–237.
168 P. Zhang, Y. Yang, Z. Huang, J. Sun, Z. Liao, J. Wang and Y. Yang, *Chem. Eng. Sci.*, 2021, **229**, 116083.
169 L. Wang, J. Liu, Y. Yan, X. Wang and T. Wang, *IEEE Trans. Instrum. Meas.*, 2017, **66**, 852–868.
170 L. Wang, Y. Yan, X. Wang, T. Wang, Q. Duan and W. Zhang, *Int. J. Greenh. Gas Control*, 2018, **68**, 269–275.
171 M. Henry, M. Tombs, M. Duta, F. Zhou, R. Mercado, F. Kenyery, J. Shen, M. Morles, C. Garcia and R. Langansan, *Flow Meas. Instrum.*, 2006, **17**, 399–413.
172 T. Green, M. Reese and M. Henry, *Meas. Control*, 2008, **41**, 205–207.
173 L. Wang, Y. Yan, X. Wang and T. Wang, *Meas. Sci. Technol.*, 2017, **28**, 035305.
174 R. May, G. Dandy and H. Maier, *Artificial Neural Networks – Methodological Advances and Biomedical Applications*, InTech, 2011.
175 A. M. Abdilahi, M. W. Mustafa, S. Y. Abujarad and M. Mustapha, *Renewable Sustainable Energy Rev.*, 2018, **81**, 3101–3110.
176 N. Mac Dowell and I. Staffell, *Int. J. Greenh. Gas Control*, 2016, **48**, 327–344.
177 W. Zhang, D. Shao, Y. Yan, S. Liu and T. Wang, *Int. J. Greenh. Gas Control*, 2018, **79**, 193–199.
178 R. T. J. Porter, M. Fairweather, M. Pourkashanian and R. M. Woolley, *Int. J. Greenh. Gas Control*, 2015, **36**, 161–174.
179 X. Cui, Y. Yan, Y. Ma, L. Ma and X. Han, *Sens. Actuators, A*, 2016, **237**, 107–118.
180 G. J. Bellante, S. L. Powell, R. L. Lawrence, K. S. Repasky and T. A. O. Dougher, *Int. J. Greenh. Gas Control*, 2013, **13**, 124–137.
181 J. L. Verkerke, D. J. Williams and E. Thoma, *Int. J. Appl. Earth Obs. Geoinf.*, 2014, **31**, 67–77.
182 Y. Chen, J. P. Guerschman, Z. Cheng and L. Guo, *Appl. Energy*, 2019, **240**, 312–326.
183 S. Sinha, R. P. de Lima, Y. Lin, A. Y. Sun, N. Symons, R. Pawar and G. Guthrie, *Int. J. Greenh. Gas Control*, 2020, **103**, 103189.
184 S. Ghanbari, Y. Al-Zaabi, G. E. Pickup, E. Mackay, F. Gozalpour and A. C. Todd, *Chem. Eng. Res. Des.*, 2006, **84**, 764–775.
185 S. Bachu, *Prog. Energy Combust. Sci.*, 2008, **34**, 254–273.
186 W. Ampomah, R. S. Balch, R. B. Grigg, B. McPherson, R. A. Will, S.-Y. Lee, Z. Dai and F. Pan, *Greenh. Gases Sci. Technol.*, 2017, **7**, 128–142.
187 G. Heddle, H. Herzog and M. Klett, *The Economics of CO2 Storage*, 2003.
188 Q. Sun, W. Ampomah, E. J. Kutsienyo, M. Appold, B. Adu-Gyamfi, Z. Dai and M. R. Soltanian, *Fuel*, 2020, **278**, 118356.
189 L. Nghiem, P. Sammon, J. Grabenstetter and H. Ohkuma, Proceedings – SPE Symposium on Improved Oil Recovery, Society of Petroleum Engineers (SPE), 2004, vol. 2004 – April.
190 L. Nghiem, V. Shrivastava, B. Kohse, M. Hassam and C. Yang, *J. Can. Pet. Technol.*, 2010, **49**, 15–22.
191 W. Ampomah, R. Balch, M. Cather, D. Rose-Coss, Z. Dai, J. Heath, T. Dewers and P. Mozley, *Energy Fuels*, 2016, **30**, 8545–8555.
192 E. J. Kutsienyo, W. Ampomah, Q. Sun, R. S. Balch, J. You, W. N. Aggrey and M. Cather, Society of Petroleum Engineers – SPE Europec Featured at 81st EAGE Conference and Exhibition 2019, Society of Petroleum Engineers, 2019.
193 F. Jiang, J. Yang, E. Boek and T. Tsuji, *Adv. Water Resour.*, 2021, **147**, 103797.
194 J. Bradshaw, S. Bachu, D. Bonijoly, R. Burruss, S. Holloway, N. P. Christensen and O. M. Mathiassen, *Int. J. Greenh. Gas Control*, 2007, **1**, 62–68.
195 S. Bachu, D. Bonijoly, J. Bradshaw, R. Burruss, S. Holloway, N. P. Christensen and O. M. Mathiassen, *Int. J. Greenh. Gas Control*, 2007, **1**, 430–443.
196 M. N. Amar and A. Jahanbani Ghahfarokhi, *J. Pet. Sci. Eng.*, 2020, **190**, 107037.
197 A. Kirch, Y. M. Celaschi, J. M. De Almeida and C. R. Miranda, *ACS Appl. Mater. Interfaces*, 2020, **12**, 15837–15843.
198 J. E. Killough, *Soc. Pet. Eng. AIME J.*, 1976, **16**, 37–48.
199 H. Ni and S. M. Benson, *Water Resour. Res.*, 2020, **56**, e2020WR027473.

200 N. A. Menad, A. Hemmati-Sarapardeh, A. Varamesh and S. Shamshirband, *J. CO2 Util.*, 2019, **33**, 83–95.
201 J. Zhang, Q. Feng, X. Zhang, C. Shu, S. Wang and K. Wu, *Energy Fuels*, 2020, **34**, 7353–7362.
202 D. Lumley, *Lead. Edge*, 2010, **29**, 150–155.
203 X. Jiang, *Appl. Energy*, 2011, **88**, 3557–3566.
204 R. Zhang, D. Vasco, T. M. Daley and W. Harbert, *Interpretation*, 2015, **3**, SM37–SM46.
205 Z. Wang, R. M. Dilmore and W. Harbert, *Int. J. Greenh. Gas Control*, 2020, **100**, 103115.
206 R. P. de Lima and Y. Lin, SEG International Exposition and Annual Meeting 2019, Society of Exploration Geophysicists, 2020, pp. 2333–2337.
207 Z. Zhong, A. Y. Sun, Q. Yang and Q. Ouyang, *J. Hydrol.*, 2019, **573**, 885–894.
208 H. Singh, *J. Nat. Gas Sci. Eng.*, 2019, **69**, 102933.
209 B. Hill, S. Hovorka and S. Melzer, *Energy Procedia*, Elsevier Ltd, 2013, vol. 37, pp. 6808–6830.
210 W. D. Gunter, S. Wong, D. B. Cheel and G. Sjostrom, *Appl. Energy*, 1998, **61**, 209–227.
211 B. Nimana, C. Canter and A. Kumar, *Appl. Energy*, 2015, **143**, 189–199.
212 Z. Dai, H. Viswanathan, R. Middleton, F. Pan, W. Ampomah, C. Yang, W. Jia, T. Xiao, S. Y. Lee, B. McPherson, R. Balch, R. Grigg and M. White, *Environ. Sci. Technol.*, 2016, **50**, 7546–7554.
213 W. Yu, H. R. Lashgari, K. Wu and K. Sepehrnoori, *Fuel*, 2015, **159**, 354–363.
214 F. Pan, B. J. McPherson, Z. Dai, W. Jia, S. Y. Lee, W. Ampomah, H. Viswanathan and R. Esser, *Int. J. Greenhouse Gas Control*, 2016, **51**, 18–28.
215 R. Balch and B. McPherson, *Society of Petroleum Engineers – SPE Western Regional Meeting*, Society of Petroleum Engineers, 2016, pp. 23–26.
216 M. L. Godec, V. A. Kuuskraa and P. Dipietro, *Energy Fuels*, 2013, **27**, 4183–4189.
217 M. K. Verma, Fundamentals of carbon dioxide-enhanced oil recovery (CO_2-EOR): a supporting document of the assessment methodology for hydrocarbon recovery using CO_2-EOR associated with carbon sequestration, Open-File Report 2015-1071, USGS Numbered Series, 2015, https://pubs.er.usgs.gov/publication/ofr20151071.
218 E. Manrique, C. Thomas, R. Ravikiran, M. Izadi, M. Lantz, J. Romero and V. Alvarado, Proceedings – SPE Symposium on Improved Oil Recovery, Society of Petroleum Engineers (SPE), 2010, vol. 2, pp. 1584–1604.
219 Y. M. Han, C. Park and J. M. Kang, *Society of Petroleum Engineers (SPE)*, 2010, pp. 14–17.
220 M. Christie, D. Eydinov, V. Demyanov, J. Talbot, D. Arnold and V. Shelkov, Society of Petroleum Engineers – SPE Reservoir Simulation Symposium 2013, Society of Petroleum Engineers, 2013, **vol. 1**, pp. 57–67.
221 A. Forooghi, A. A. Hamouda and T. Eilertsen, All Days, SPE, 2009.
222 J. You, W. Ampomah, E. J. Kutsienyo, Q. Sun, R. S. Balch, W. N. Aggrey and M. Cather, Society of Petroleum Engineers – SPE Europec Featured at 81st EAGE Conference and Exhibition 2019, Society of Petroleum Engineers, 2019.
223 G. Vida, M. D. Shahab and M. Mohammad, *Fluids*, 2019, **4**, 85.
224 E. Artun, *J. Pet. Sci. Eng.*, 2020, **195**, 107768.
225 S. Amini, S. Mohaghegh, R. Gaskari and G. Bromhal, Society of Petroleum Engineers Western Regional Meeting 2012, Society of Petroleum Engineers, 2012, pp. 557–566.
226 S. Amini, S. D. Mohaghegh, R. Gaskari and G. S. Bromhal, Society of Petroleum Engineers – SPE Intelligent Energy International 2014, Society of Petroleum Engineers (SPE), 2014, pp. 781–787.
227 S. Amini and S. Mohaghegh, *Fluids*, 2019, **4**, 126.
228 H. Xiong, C. Kim and J. Fu, Proceedings – SPE Symposium on Improved Oil Recovery, Society of Petroleum Engineers (SPE), 2020, vol. 2020 – August.
229 S. Raha Moosavi, D. A. Wood and A. Samadani, *Comput. Res. Prog. Appl. Sci. Eng. (CRPASE)*, 2020, **6**, 1–8.
230 B. Chen and R. J. Pawar, *Energy*, 2019, **183**, 291–304.
231 A. Y. Sun, *Appl. Energy*, 2020, **278**, 115660.
232 N. A. Menad and Z. Noureddine, *J. Taiwan Inst. Chem. Eng.*, 2019, **99**, 154–165.
233 Y. Zhang and N. V. Sahinidis, *Ind. Eng. Chem. Res.*, 2013, **52**, 3121–3132.
234 J. You, W. Ampomah, Q. Sun, E. J. Kutsienyo, R. S. Balch, Z. Dai, M. Cather and X. Zhang, *J. Clean. Prod.*, 2020, **260**, 120866.
235 J. You, W. Ampomah and Q. Sun, *Fuel*, 2020, **264**, 116758.
236 J. You, W. Ampomah, Q. Sun, E. J. Kutsienyo, R. S. Balch and M. Cather, Proceedings – SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers (SPE), 2019, vol. 2019-September, p. 196182.
237 A. Almasov, M. Onur and A. C. Reynolds, Proceedings – SPE Symposium on Improved Oil Recovery, Society of Petroleum Engineers (SPE), 2020, vol. 2020 – August.
238 M. N. Amar, N. Zeraibi and A. Jahanbani Ghahfarokhi, *Greenhouse Gases: Sci. Technol.*, 2020, **10**, 613–630.
239 A. Nwachukwu, H. Jeong, A. Sun, M. Pyrcz and L. W. Lake, Proceedings – SPE Symposium on Improved Oil Recovery, Society of Petroleum Engineers (SPE), 2018, vol. 2018 – April.
240 W. Ampomah, R. S. Balch, M. Cather, R. Will, D. Gunda, Z. Dai and M. R. Soltanian, *Appl. Energy*, 2017, **195**, 80–92.
241 W. Ampomah, R. Balch, R. B. Grigg, M. Cather, E. Gragg, R. A. Will, M. White, N. Moodie and Z. Dai, *Geomech. Geophys. Geo-Energy Geo-Resour.*, 2017, **3**, 245–263.
242 G. S. Bromhal, J. Birkholzer, S. D. Mohaghegh, N. Sahinidis, H. Wainwright, Y. Zhang, S. Amini, V. Gholami, Y. Zhang and A. Shahkarami, *Energy Procedia*, 2014, **63**, 3425–3431.
243 A. Nwachukwu, H. Jeong, M. Pyrcz and L. W. Lake, *J. Pet. Sci. Eng.*, 2018, **163**, 463–475.
244 H. Yan, J. Zhang, N. Zhou and M. Li, *Sci. Total Environ*, 2020, **711**, 135029.
245 Q. Feng, J. Wang, J. Zhang and X. Zhang, *J. Phys. Conf. Ser.*, 2021, **1813**, 012023, DOI: 10.1088/1742-6596/1813/1/012023.

246 M. Meng, Z. Qiu, R. Zhong, Z. Liu, Y. Liu and P. Chen, *Chem. Eng. J.*, 2019, **368**, 847–864.
247 H. Yan, J. Zhang, N. Zhou, B. Li and Y. Wang, *Eng. Fract. Mech.*, 2021, **249**, 107750.
248 R. Zhong, R. Johnson Jr, Z. Chen and N. Chand, *APPEA J.*, 2019, **59**, 319.
249 L. K. Sharma, V. Vishal and T. N. Singh, *J. Nat. Gas Sci. Eng.*, 2017, **42**, 216–225.
250 H. Yan, J. Zhang, S. S. Rahman, N. Zhou and Y. Suo, *Sci. Total Environ.*, 2020, **705**, 135941.
251 Z. Guo, J. Zhao, Z. You, Y. Li, S. Zhang and Y. Chen, *Energy*, 2021, **230**, 120847.
252 D. Kang, X. Wang, X. Zheng and Y. P. Zhao, *Fuel*, 2021, **290**, 120006.
253 M. Meng, R. Zhong and Z. Wei, *Fuel*, 2020, **278**, 118358.
254 M. Nait Amar, A. Larestani, Q. Lv, T. Zhou and A. Hemmati-Sarapardeh, *J. Pet. Sci. Eng.*, 2021, 109226.
255 A. Bemani, A. Baghban, A. H. Mohammadi and P. Ø. Andersen, *J. Nat. Gas Sci. Eng.*, 2020, **76**, 103204, DOI: 10.1016/j.jngse.2020.103204.
256 L. Wang, M. Liu and A. Altazhanov, *et al.*, Data driven machine learning models for shale gas adsorption estimation, Paper SPE 200621 Presented at SPE Europe featured at 82nd EAGE Conference and Exhibition, Amsterdam, Netherlands, 8–11 December, 2020.
257 Z.-Z. Yang, L.-N. He, Y.-N. Zhao, B. Li and B. Yu, *Energy Environ. Sci.*, 2011, **4**, 3971–3975.
258 S. Kar, A. Goeppert and G. K. S. Prakash, *Acc. Chem. Res.*, 2019, **52**, 2892–2903.
259 V. W. Y. Tam, A. Butera, K. N. Le and W. Li, *Constr. Build. Mater.*, 2020, **250**, 118903.
260 Z. Zhang, S. Y. Pan, H. Li, J. Cai, A. G. Olabi, E. J. Anthony and V. Manovic, *Renewable Sustainable Energy Rev.*, 2020, **125**, 109799.
261 K. Tran and Z. W. Ulissi, *Nat. Catal.*, 2018, **1**, 696–703.
262 E. Boutin, L. Merakeb, B. Ma, B. Boudy, M. Wang, J. Bonin, E. Anxolabéhère-Mallart and M. Robert, *Chem. Soc. Rev.*, 2020, **49**, 5772–5809.
263 Z. W. Ulissi, M. T. Tang, J. Xiao, X. Liu, D. A. Torelli, M. Karamad, K. Cummins, C. Hahn, N. S. Lewis, T. F. Jaramillo, K. Chan and J. K. Nørskov, *ACS Catal.*, 2017, **7**, 6600–6608.
264 M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C. T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C. S. Tan, M. Askerka, F. Che, M. Liu, A. Seifitokaldani, Y. Pang, S. C. Lo, A. Ip, Z. Ulissi and E. H. Sargent, *Nature*, 2020, **581**, 178–183.
265 X. Ma, Z. Li, L. E. K. Achenie and H. Xin, *J. Phys. Chem. Lett.*, 2015, **6**, 3528–3533.
266 Z. Li, X. Ma and H. Xin, *Catal. Today*, 2017, **280**, 232–238.
267 D. Wu, J. Zhang, M. J. Cheng, Q. Lu and H. Zhang, *J. Phys. Chem. C*, 2021, **125**, 15363–15372.
268 X. Wan, Z. Zhang, H. Niu, Y. Yin, C. Kuai, J. Wang, C. Shao and Y. Guo, *J. Phys. Chem. Lett.*, 2021, **12**, 6111–6118.
269 A. Chen, X. Zhang, L. Chen, S. Yao and Z. Zhou, *J. Phys. Chem. C*, 2020, **124**, 22471–22478.
270 Y. Guo, X. He, Y. Su, Y. Dai, M. Xie, S. Yang, J. Chen, K. Wang, D. Zhou and C. Wang, *J. Am. Chem. Soc.*, 2021, **143**, 5755–5762.
271 S. Li, Y. Zhang, Y. Hu, B. Wang, S. Sun, X. Yang and H. He, *J. Mater.*, 2021, **7**, 1029–1038.
272 Y. A. Lim, M. N. Chong, S. C. Foo and I. M. S. K. Ilankoon, *Renewable Sustainable Energy Rev.*, 2021, **137**, 110579.
273 A. Coşgun, M. E. Günay and R. Yıldırım, *Renewable Energy*, 2021, **163**, 1299–1317.
274 W. Z. Taffese, E. Sistonen and J. Puttonen, *Constr. Build. Mater.*, 2015, **100**, 70–82.
275 Y. Song, K. Yang, J. Chen, K. Wang, G. Sant and M. Bauchy, *ACS Sustainable Chem. Eng.*, 2021, **9**, 2639–2650.