Application of Machine Learning in Carbon Capture and Storage (CCS)

M. Taha Usman

International School Lahore

Abstract

The use of machine learning (ML) in carbon capture and storage (CCS) is an evolving field, with many potential technologies in the pipeline for implementation. This paper considers the most prevalent ones, including ANN, DT, SVM(R), XGBoost, and clustering, with their unique toolsets, and how they can be implemented in hybrid models to ensure effective CCS. It considers how these technologies are trained, and how they can be implemented in various parts of the CCS process, including storage site selection, real-time monitoring and optimisation, leakage detection, predictive maintenance, and enhancing CO_2 absorption materials. The investigation of the XGBoost algorithm in this study has confirmed ML's effectivity, whilst identifying areas of further improvement which can be worked upon to enhance the model's accuracy (root squared, or R^2, score).

1. Introduction

The escalating threat of climate change has underscored the urgent need for effective strategies to mitigate greenhouse gas emissions, with carbon dioxide (CO_2) being a primary target due to its substantial contribution to global warming (Intergovernmental Panel on Climate Change (IPCC), 2021)[1]. Carbon Capture and Storage (CCS) has emerged as a pivotal technology in this mitigation strategy, aimed at reducing atmospheric CO_2 concentrations by capturing carbon emissions from industrial processes and securely storing them underground (Adjiman, et al., 2018)[2] (Shreyash, et al., 2021)[3]. Despite significant advancements in CCS technology, challenges related to efficiency, cost, and operational management persist, necessitating innovative approaches to enhance its effectiveness and scalability.

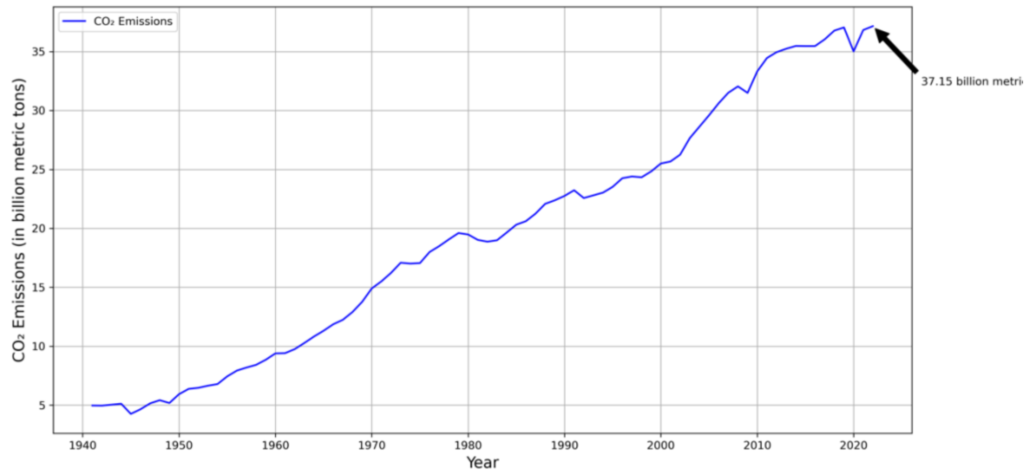


FIGURE 1 - Carbon Emissions mapped out over an approximately 80-year period.

Recent developments in machine learning (ML) offer promising avenues to address these challenges. Machine learning, a subset of artificial intelligence, encompasses various techniques and algorithms that enable systems to learn from data and improve performance over time without explicit programming. By leveraging vast amounts of data and advanced computational capabilities, ML can optimize various aspects of CCS operations, from capture and transportation to storage and monitoring.

The integration of ML into CCS processes holds the potential to revolutionize the field in several ways (Yao, Yu, Zhang, & Xu, 2023)[4]. For instance, ML algorithms can improve the efficiency of capture technologies by predicting optimal operating conditions and identifying anomalies. In the transportation phase, ML can enhance pipeline monitoring and predictive maintenance, reducing the risk of leaks and failures. Furthermore, ML models can play a crucial role in storage site characterization and monitoring by analysing geological data to predict the behaviour of CO_2 in storage sites, ensuring long-term storage security.

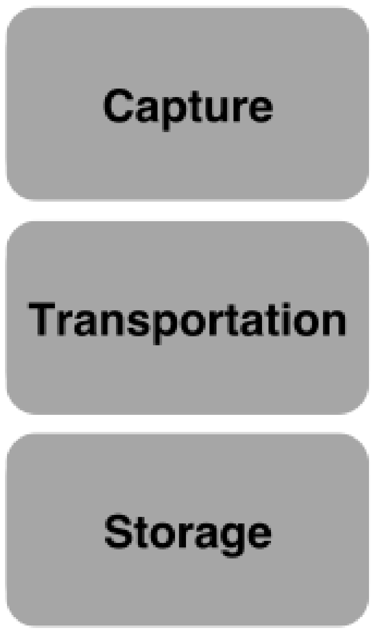
Despite its potential, the application of ML in CCS is still in its nascent stages, and several hurdles remain, including data quality, model interpretability, and integration with existing CCS infrastructure. This paper aims to explore the current state of ML applications in CCS, highlighting key advancements, ongoing research, and practical implementations. The paper is structured as follows: Section 2 provides an overview of CCS technology and its operational challenges. Section 3 delves into various ML techniques and their applications within different phases of CCS. Section 4 discusses case studies and real-world examples where ML has been successfully applied to CCS. Section 5 identifies ongoing research directions and potential future developments in this interdisciplinary field. Finally, Section 6 concludes with a summary of findings and recommendations for future research.

By examining the intersection of ML and CCS, this paper contributes to a deeper understanding of how emerging technologies can enhance climate change mitigation efforts and offers insights into potential pathways for future innovations in the field. It is determined that ML offers great help in analysing vast parameters and modelling the long-term effectiveness of a geological structure in storing carbon dioxide.

2. Carbon Capture and Storage (CCS)

Carbon Capture and Storage (CCS) is a critical technology designed to reduce carbon dioxide (CO_2) emissions from industrial processes and power generation, thereby mitigating climate change (Budinis, Krevor, Mac Dowell, Brandon, & Hawkes, 2018)[5]. This section provides an in-depth overview of CCS technology, including its main components, processes, and operational challenges.

CCS involves three main stages: capture, transportation, and storage. Each stage is integral to the overall efficacy of the technology. The primary focus of this study is how CCS can be effectively implemented to offset the dangerously large number of emissions being released



daily. However, understanding the inherent challenges of CCS at present is essential for its effective implementation. Defining the term first, CCS refers to a vast range of processes for removing CO_2 from a point source (such as industrial flue gas) or the atmosphere via direct air capture. Once the CO_2 has been captured, it can either be put into permanent storage, usually underground, or utilized to manufacture valuable products like fuels or specialty chemicals, as part of carbon utilization (Boot-Handford, et al., 2014)[6] (Rahimi, Moosavi, Smit, & Hatton, 2021)[7].

FIGURE 2 - The three stages of CCS

In the capture phase, CO_2 is extracted from industrial emissions or power plant exhaust gases. There are three primary capture methods: pre-combustion, post-combustion, and oxy-fuel combustion. Pre-combustion capture converts fossil fuels into a mixture of hydrogen and CO_2 before combustion, allowing for the CO_2 to be separated and captured. This method is efficient but requires significant modifications to existing power plants. Post-combustion capture, the most commonly used method, involves capturing CO_2 from flue gases produced after the combustion of fossil fuels. This approach can be retrofitted to existing power plants, making it more versatile. Oxy-fuel combustion burns fossil fuels in the presence of pure oxygen, resulting in a flue gas that is mainly water vapour and CO_2, which can then be easily separated.

Advanced CCS technologies primarily use an absorption route, where an absorbent agent captures CO_2 from a mixed gas stream (Rahimi, Moosavi, Smit, & Hatton, 2021)[7]. There is a subsequent thermal-stripping process where pure CO_2 is released, and the absorbent agent is regenerated. Most thermal-based capture systems use an amine, such as diethanolamine (DEA), methyl diethanolamine (MDEA), or monoethanolamine (MEA). However, adsorbent-based processes, which are more energy-efficient and versatile, also exist. In these processes, CO_2 is selectively adsorbed on the surface (or within the matrix) of an adsorbent substrate, which can subsequently be regenerated by a pressure or thermal swing. Typical adsorbents include zeolite, activated carbon, metal oxides, and silica gel.

Delving deeper into the different technologies that exist, membrane technology utilizes the Knudsen diffusion principle, whereby CO_2 dissolves in the membrane and diffuses at a rate proportional to its partial pressure gradient (Wilberforce, Baroutaji, Soudan, Al-Alami, & Olabi)[8]. This method is especially useful for removing CO_2 from natural gas or where its pressure is high. However, capturing CO_2 from flue gas, due to its lesser quantity, poses a challenge because of the greater energy requirement to achieve the necessary carbon capture ratio. Moreover, cryogenic separation involves various compression applications at ambient temperature and pressure to separate the gas, making it viable for producing liquid CO_2 and high-concentration CO_2 capture.

In the transportation phase, CO_2 must be transported to storage sites. Transportation is typically carried out through pipelines, which are considered the most economical and efficient method for moving large volumes of CO_2 over long distances. In some cases, CO_2 can also be transported by ships, particularly when pipelines are not feasible due to geographical constraints. The integrity and safety of CO_2 transport are paramount to prevent leaks and ensure that the gas reaches its storage site without causing environmental harm.

The final stage of CCS is storage, where CO_2 is injected into deep geological formations for long-term isolation from the atmosphere. Suitable storage sites include depleted oil and gas fields, deep saline aquifers, and unmineable coal seams. These geological formations are chosen based on their capacity to securely contain CO_2 and their impermeability, which prevents CO_2 from escaping to the surface. The process of injecting CO_2 into these formations is known as geological sequestration. Extensive site characterization and monitoring are required to ensure the long-term stability and security of stored CO_2. This involves detailed geological surveys, risk assessments, and the implementation of monitoring technologies to detect any potential leaks or changes in the storage site.

Despite its potential, CCS technology faces several significant challenges. One of the primary challenges is the high cost associated with the capture phase, which can account for up to 70% of the total cost of CCS. The energy-intensive nature of CO_2 capture processes reduces the overall efficiency of power plants and industrial operations, thereby increasing operational costs. Additionally, the development and maintenance of CO_2 transportation infrastructure, such as pipelines, require substantial financial investment. The long-term monitoring and verification of storage sites also add to the overall cost and complexity of CCS projects (Boot-Handford, et al., 2014)[6] (Pires, Martins, Alvim-Ferraz, & Simões, 2011)[8] (Anderson & Newell, 2004)[9] (Gibbins & Chalmers, 2008)[10] (Adjiman, et al., 2018)[2].

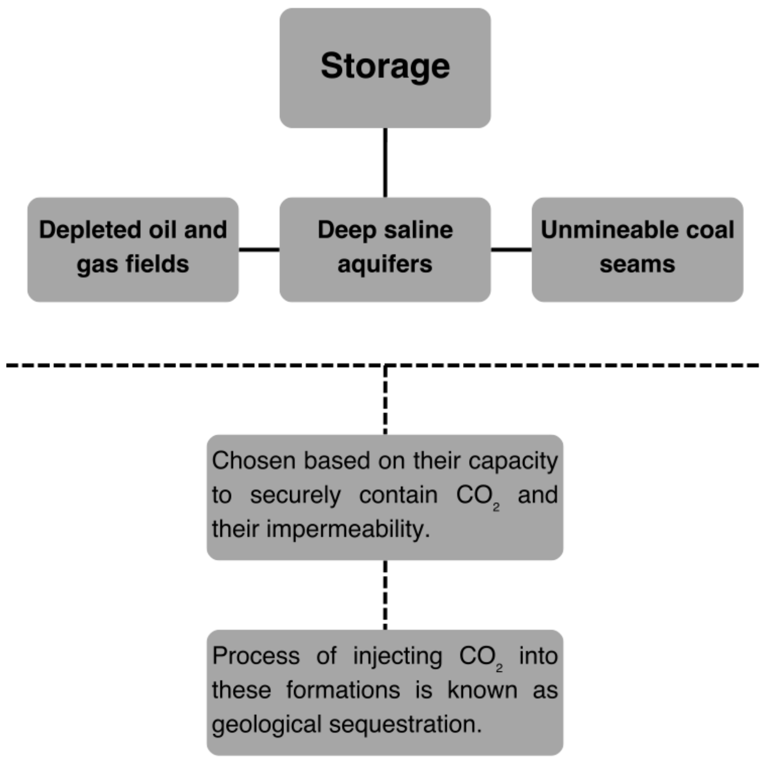


FIGURE 3 - The carbon storage process

Another major challenge is the regulatory and policy framework surrounding CCS. Clear and supportive policies are essential to incentivize investment in CCS technologies and ensure compliance with safety and environmental standards. The lack of a comprehensive legal framework can hinder the development and deployment of CCS projects. Moreover, public perception and acceptance of CCS can be a barrier. Concerns about the safety of CO_2 storage, potential environmental impacts, and the perception that CCS may prolong the use of fossil fuels can lead to resistance from local communities and stakeholders.

Addressing these challenges requires a multifaceted approach. Continued research and development are crucial to advancing CCS technologies and reducing costs. Innovations in materials, capture processes, and monitoring techniques can enhance the efficiency and reliability of CCS. Additionally, government policies and incentives, such as carbon pricing and subsidies, can stimulate investment and adoption of CCS. Public engagement and education are also vital to address misconceptions and build trust in CCS as a viable climate mitigation strategy.

Carbon Capture and Storage is a pivotal technology in the fight against climate change, offering a means to significantly reduce CO_2 emissions from industrial sources. While the technology has made considerable progress, overcoming the challenges of cost, regulatory frameworks, and public perception is essential for its widespread implementation. By addressing these issues and leveraging advancements in technology, including the integration of machine learning, CCS can play a crucial role in achieving global climate goals and transitioning to a low-carbon future. The next section will explore the application of machine learning in CCS, detailing how these advanced computational techniques can address some of the current challenges and enhance the overall efficiency and effectiveness of CCS processes.

3. Machine learning techniques in CCS

Machine learning offers a variety of solutions to these issues, with each unique model offering its own set of features to tackle a different problem.

3.1 Artificial Neural Networks (ANNs)

One of the core technologies is Artificial Neural Networks (ANNs). An ANN consists of 3 layers: input, hidden (where the processing occurs), and output. These are based on a perceptron algorithm, which imitates human neurons (Yan, et al., 2021)[11]. ANNs are commonly used to project CO_2 solubility, viscosity, saturation and density, especially when the carbon interacts with multicomponent gas-liquid mixtures, which ensures its long-term safety and effective storage (such as to predict the storage efficiency in a saline aquifer), along with optimising the conditions for enhanced oil recovery (EOR), which is an effective way to repurpose the CO_2 (Yao, Yu, Zhang, & Xu, 2023)[4] (Vaziri & Sedaee, 2023)[12]. Saline aquifers are a poignant consideration as they are one of the most effective means of long-term carbon storage. Later in the paper, a study by Song et al (2020)[13] is considered, which used an ANN to create a synthetic model of saline aquifers to map out their long-term carbon impermeability.

ANN's use further extends to being able to simulate subsurface characteristics at the reservoir scale from permeability parameters and core porosity in the event of incomplete and insufficient site data, facilitating the process of finding optimal storage sites.

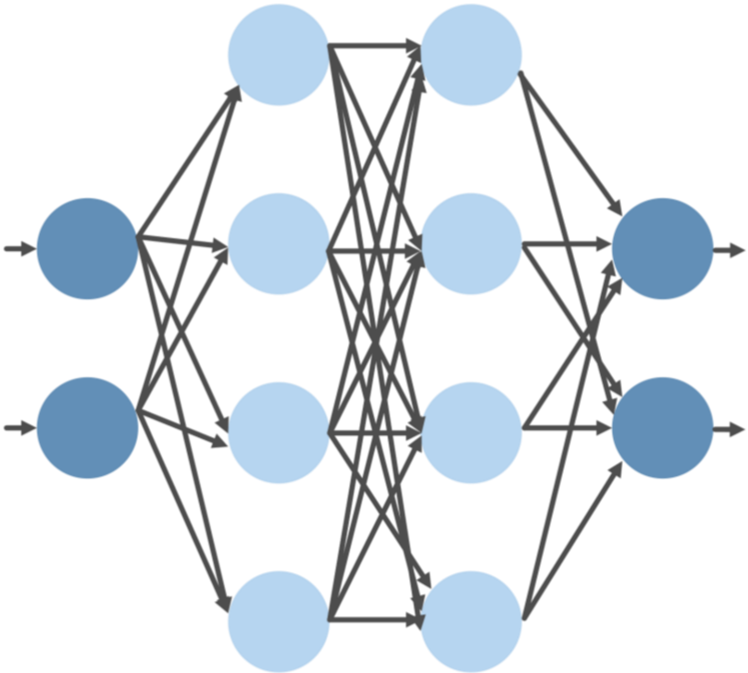


FIGURE 4 - ANN

3.2 Deep Neural Networks (DNN)

For even more computationally intensive applications of ANN, DNNs are very robust. A subset of ANN that is made up of a greater number of deeper layers, the deep architecture of DNN allows the networks to automatically learn hierarchical representations of data and capture its complex patterns and features (Nassabeh, You, Keshavarz, & Iglauer, 2024)[13]. For this, it utilises a process called forward propagation (in which data passes through the hidden layers), and the complementary processes of back-propagation to reduce the difference between the predicted output and the actual target values; this entails the adjusting of weights and biases using optimisation algorithms like gradient descent (Nassabeh, You, Keshavarz, & Iglauer, 2024)[13]. This results in DNNs being able to effectively capture both local and global patterns in CCS research, particularly in the case of non-linear relationships, and dealing with vast and complex datasets made up of environmental and geological factors.

3.3 Random Forest (RF) and Decision Tree (DT)

Moving on to a more distinct technique, RF shows potential in enacting a feasibility assessment of CCS along with other technologies, such as biomass conversion and hydrothermal treatment (HTT) (Yao, Yu, Zhang, & Xu, 2023)[11]. RF is a subset of the Decision Tree (DT) model, and these, coupled, are effective in risk assessment, decision analysis and estimating the success probability of CCS. DT involves a process called feature selection, where the model chooses a feature that best separates data based on a chosen metric, establishing a root node. The process repeats such that an increasing number of internal nodes are formed to further classify the data until leaf (end) nodes are formed to give final predictions (Song & Lu, 2015)[15]. RF builds on DT by creating an ensemble of DTs, each being trained on a different subset of data, and selected through the processes of bootstrap sampling. RF brings in additional randomness by sampling subsets of the training data and features during tree construction, which reduces overfitting and increases the accuracy (Nassabeh, You, Keshavarz, & Iglauer, 2024)[14]. Each tree votes on predictions in the forest, with the final result being determined by the majority vote (in the case of classification) or averaging (in the case of regression) (Nassabeh, You, Keshavarz, & Iglauer, 2024)[14]. This algorithm holds the benefits of being able to provide strong predictions and determine feature importance for each parameter input. For this reason, RF is optimal for processes such as CCS site screening, because of its ability to map out complex relationships between economic, geological, and environmental parameters (Nassabeh, You, Keshavarz, & Iglauer, 2024)[14]. Its feature importance analysis helps it with decision-making by highlighting key factors, which allow for the evaluation of the success probability of a CCS project; added to this is its ease of use and scalability, which enable it to deal with large and diverse datasets during the screening research process (Nassabeh, You, Keshavarz, & Iglauer, 2024)[14] (Yao, Yu, Zhang, & Xu, 2023)[11]. Furthermore, the model is fail-safe against overfitting (which would entail failing to accurately perform on unseen data) due to its ensemble property (Nassabeh, You, Keshavarz, & Iglauer, 2024)[14].

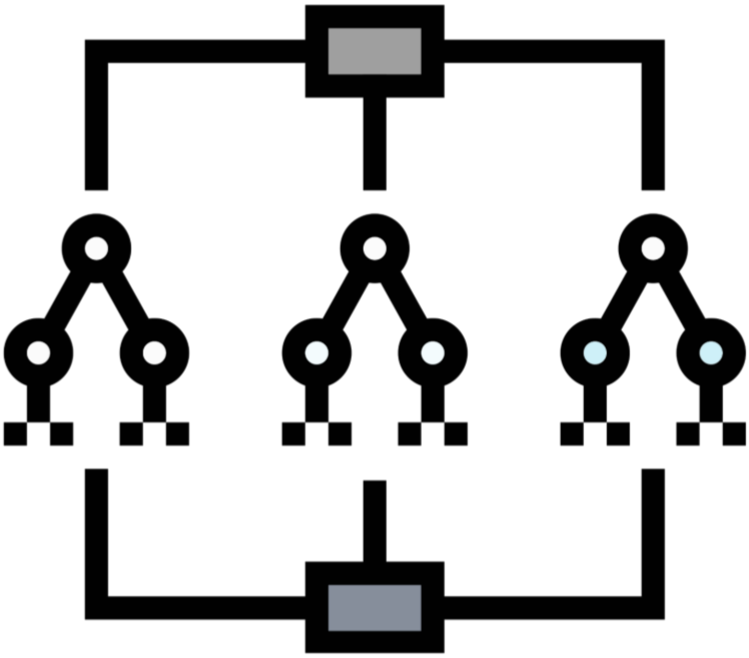


FIGURE 5 - RF

3.4 Extreme Gradient Boosting (XGBoost)

A related ML model is XGBoost, which utilises gradient boosting, a technique used for regression and classification tasks that sequentially builds an ensemble of trees, with each tree correcting the errors of the previous trees. Also utilising ensemble learning, XGBoost is optimal for screening research in CCS site screening and selection, as well as being able to capture nonlinear relationships, offer feature importance insights, and deal with missing data (Nassabeh, You, Keshavarz, & Iglauer, 2024)[14]. With its regularisation techniques, it can provide flexibility and optimal performance to safeguard against overfitting, allow efficient training, and provide a wide range of tunable hyperparameters (Nassabeh, You, Keshavarz, & Iglauer, 2024)[14].

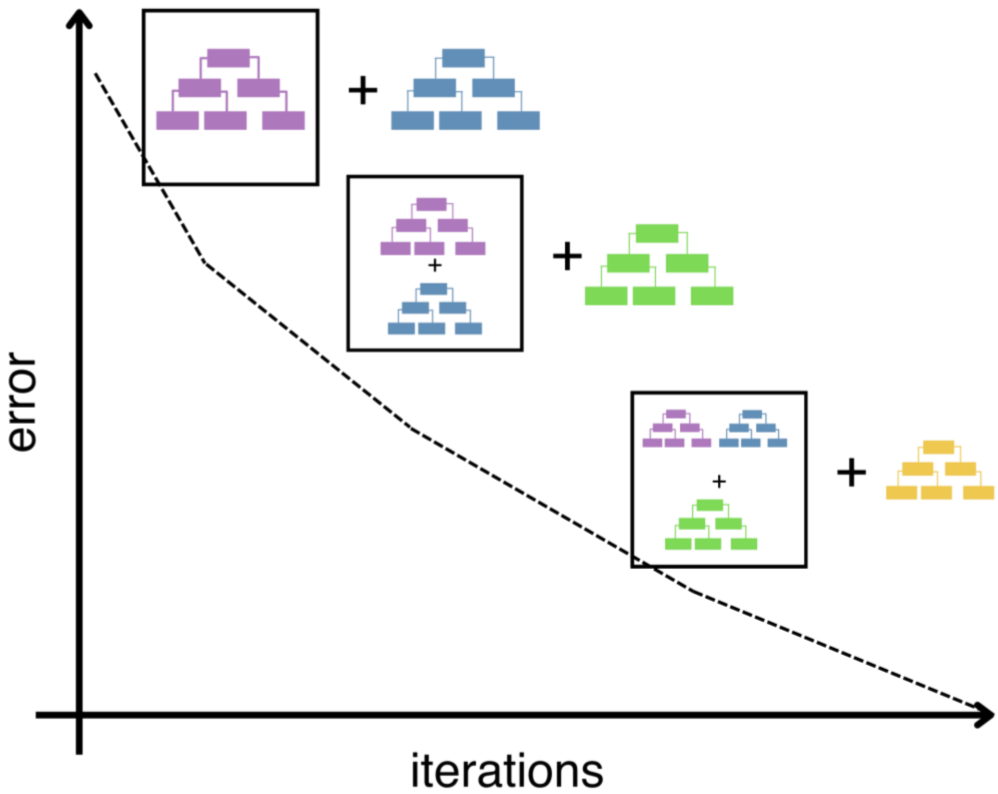


FIGURE 6 - Gradient Boosting

3.5 Support Vector Machine (SVM) and Support Vector Regressor (SVR)

SVM is a supervised machine learning algorithm that is primarily used for classification tasks, with SVR being the regression counterpart. Its main strength lies in its ability to deal with high-dimensional data, as well as to create robust classifiers with good generalisation capabilities. SVM represents training samples as vectors called support vectors in a space mapped such that samples from the separate categories are divided by a clear gap, which is as wide as possible

(Balabin & Lomakina, 2010)[16]. SVM uses a hyperplane to separate data points of different classes, with each support vector having its margin (distance) from the hyperplane, and the model aims to maximise this margin to improve generalisation and avoid excessive variance. Similarly, SVR finds a linear relation between the regressors (input variables, X) and the dependent variables (y), and maps out the cost function, which is minimised to arrive at the best regression model (Balabin & Lomakina, 2010)[16]. The aim is to use the cost function to minimise both the coefficients' size and the prediction errors (function smoothness and accuracy). Prediction errors are penalised linearly, except those that have a deviation below a certain value, ε, which is defined by Vapnik's ε-insensitive loss function. SVR's use in CCS lies in its ability to identify dominant risk factors, as well as to perform sensitivity analysis on parameter uncertainty, which aids the solution of multi-objective optimisation problems (Yao, Yu, Zhang, & Xu, 2023)[11]. Moreover, SVM is seen as a reliable method for estimating minimum miscibility pressure (MMP) in CO_2-EOR, which is important in project design and reservoir screening (Yao, Yu, Zhang, & Xu, 2023)[11]. Furthermore, LS-SVM is superior to various methods in predicting physical parameters like viscosity, the solubility of CO_2 in the saline layer, and the thermal conductivity of CO_2.

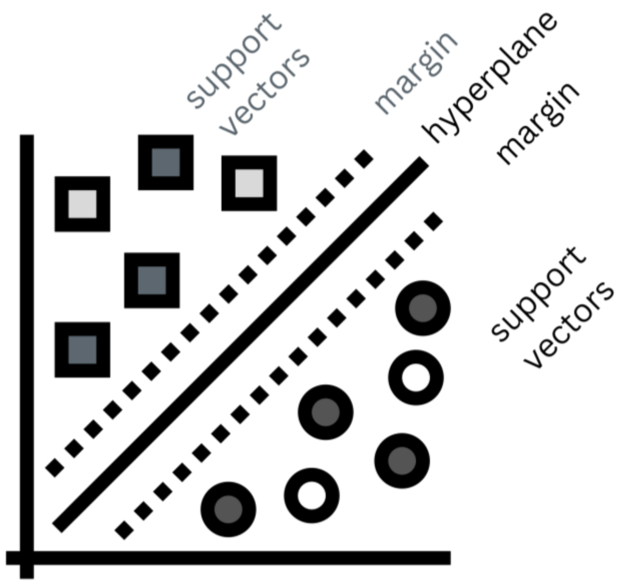


FIGURE 7 - Support Vector Model

3.6 Deep Learning (DL)

On the other hand, DL offers the ability to achieve better fitting with less data, owing to its construction as a neural network with multiple hidden layers (Yao, Yu, Zhang, & Xu, 2023)[11]. This also enables it to discover complex structures in high-dimensional data, whereby it forms a more abstract representation of higher-level attributes by combining lower-level features. One DL technique that has been implemented in CCS is a Convolutional Neural Network (CNN), a technique that is highly effective in reservoir property prediction and CO_2 plume migration tracking model development, due to its powerful image processing capability (Yao, Yu, Zhang, & Xu, 2023)[11]. Zhong et al (2019) proposed a DL model for monitoring pressure anomalies in real-time to ensure reservoir safety and integrity which uses CNN for spatial pattern mining and a technique called Long Short-Term Memory (LSTM) for temporal pattern recognition. LSTM is a subset of Recurrent Neural Network (RNN), a DL model that can memorise and store previous information and which is specialised in sequential data processing such as text and speech, with LSTM being used for long-term data storage.

3.7 Clustering

Clustering algorithms are a significant consideration for CCS, with one of the most effective ones being K-means Clustering. This unsupervised machine learning algorithm groups an unlabelled dataset into different clusters. It aims to ensure that the distance of data points within each cluster is as small as possible and that the distance between clusters is as large as possible

(Hengrui & Olegovna)[17]. Clustering algorithms are used in the preliminary stages of CCS projects to classify geological formations, as well as to identify potential storage sites based on similar characteristics, allowing for the efficient allocation of resources (Pires et al., 2011)[8] (Budinis et al., 2018)[4].

3.8 Hybrid Models

A combination of these models would be utilised in various situations for different parts of the CCS process. An example of this would be to integrate ANNs with SVMs or RFs, which improves the prediction of CO_2 capture efficiency and storage security. Broadly, combining the pattern recognition capabilities of ANNs with the classification strength of SVMs, or the ensemble approach of RFs (Gibbins & Chalmers, 2008)[10] helps provide more accurate and reliable predictions.

4. **Case Studies of ML in CCS**

4.1 Monitoring and Optimisation, including Leakage Detection

This section aims to holistically consider the various ML techniques described above in the context of how they facilitate different parts of the CCS process. Firstly, for monitoring and optimisation, one of the most effective techniques is CNN, whose convolutional layers can manage local spatial features, as well as carry out image processing. This can be implemented in saline aquifers, where changes in salinity, CO_2 saturation and mineral types can be detected by electromagnetic monitoring to ensure they do not affect complex resistivity. Similarly, using

CNN, the distribution area and real-time location of CO_2 plum in deep reservoirs can be determined using the electromagnetic monitoring data on the ground.

Similarly, a Generative Adversarial Network (GAN) can be used to map a relationship between reservoir permeability and CO_2 plume migration based on 4D seismic data (Zhong et al, 2019). Zhong et al (2019) event presented a cyclic GAN to deduce the mapping relationship between seismic impedance and CO_2 saturation in reservoirs, which can monitor leakage.

4.2 Storage Site Selection

For storage site selection, it is important to consider various parameters, such as the trapping mechanism. Whilst DT and RF are suited for this, a case study mentioned earlier demonstrates the great potential of ANN for this stage. For this, ANNs were considered in a study by Song et al (2020)[12], where an ANN-GCS (geological CO_2 sequestration) model was tested at the Pohang Basin near Pohang, South Korea. The saline aquifer here contained coarse-grained conglomerate and sandstone at depths exceeding 740 m, and coarse-grained rocks covered by a thick layer of mudstone more than 700 m in thickness, which prevents the leakage of CO_2. Geologic and petrophysical data (such as for hydrostatic pressure, geothermal gradient, and salinity) was used to develop a geological model of the Pohang Basin, which was used for dynamic CO_2 modelling, and to create a synthetic model of saline to generate residual trapping index (RTI) and solubility trapping index (STI) datasets as input and output. The model was subsequently developed with structural designs of optimal neurons and networks, the optimal stability and processes of the neural network being determined by validating the model through data generated from the field application model of the Pohang Basin. To generate the datasets, simulations of a synthetic model were performed using a GEM reservoir simulation package, which made use of the material balance equations for gas, water, and mineral components in the synthetic reservoir model. In the resultant simulation, CO_2 was injected into the aquifer at a rate of 80 tons per day over ten years, followed by a shut-in phase for 290 years, which allowed the team to monitor the long-term behaviour of the stored CO_2. This allowed them to assess the distribution of CO_2 saturation, pore pressure, and the effectiveness of various trapping mechanisms, including residual trapping, solubility trapping, structural trapping, and mineral trapping.

4.3 Case study - using Machine Learning to Validate Carbon Containment in the Illinois Basin

To further investigate ML's effectiveness, this study considers the use of the XGBoost algorithm to validate carbon containment in the Illinois Basin. The raw dataset contained various parameters, all intended to be used to determine the injection rate delta. The Illinois Basin - Decatur Project has been established to demonstrate the capacity, injectivity and containment of carbon storage in the Mount Simon Sandstone, a main carbon storage resource in the Illinois Basin in the Midwest United States (US) (Machine Learning Challenge – Using AI to Validate Carbon Containment in the Illinois Basin, n.d.). A distributed temperature sensor (DTS) fibre optic cable was installed in the tubing and extended to a depth of 6,326 feet to monitor temperature changes, taking readings every 1.624 feet every 5 seconds. The feature variables were plotted as histograms (figure 8) and boxplots (figure 9) to evaluate the data distribution and identify any outliers.

FIGURE 8 - Histogram representation of the data

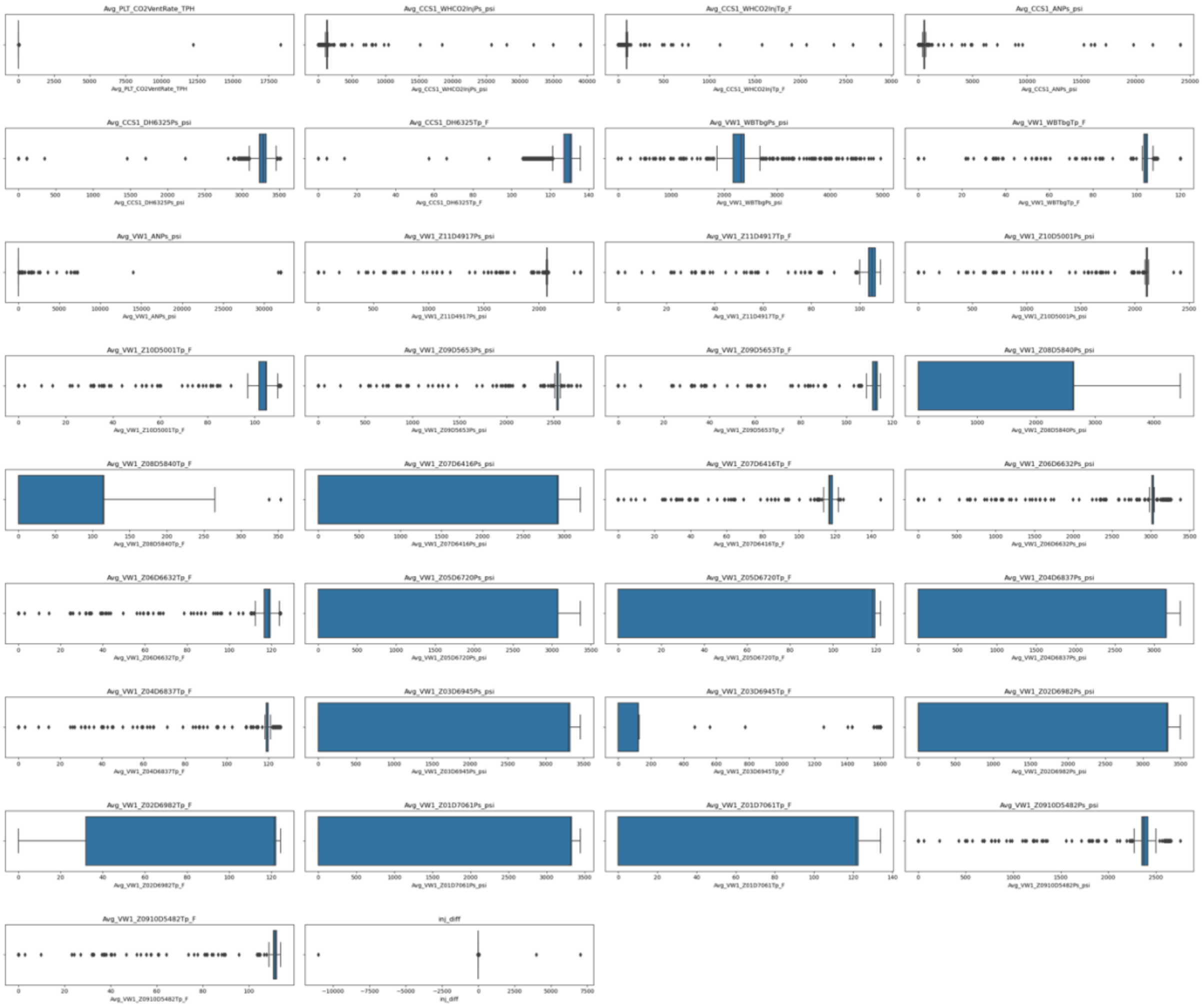


FIGURE 9 - Boxplot representation of the data

To reduce the impact of the outliers, outliers that exceeded the upper and lower interquartile limit (greater than 1.5 times the interquartile range (IQR) at both ends) were removed. The data was subsequently plotted on a correlation matrix to determine which features affect the target variable the most; an example of a high positive correlation feature is the CO_2 vent rate.

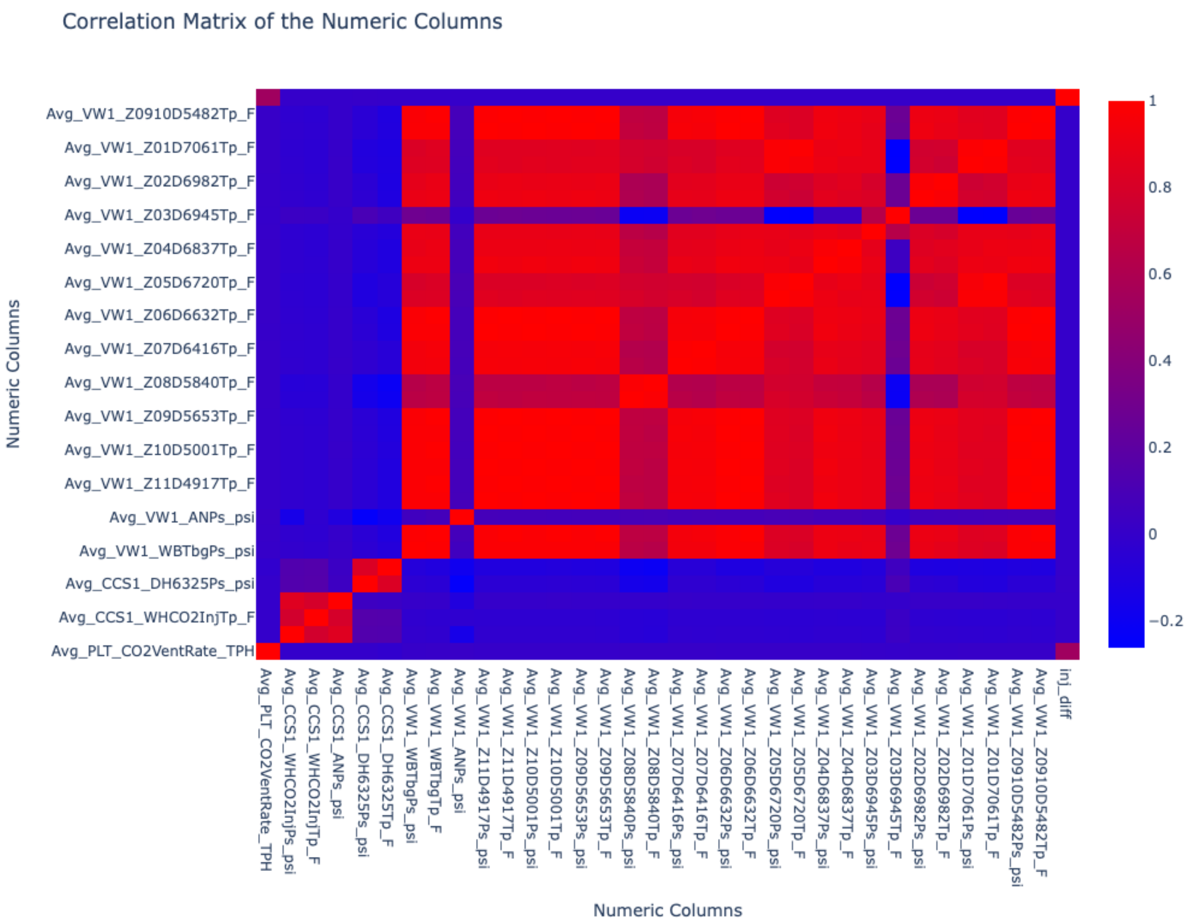


FIGURE 10 - Correlation matrix of numeric columns (feature set)

This gave a clearer understanding of which features are the most important to consider when collecting data on carbon storage structures.

Following this preliminary assessment, the data was trained, the sklearn library being used extensively to split the data into training and testing variables using train_test_split, and StandardScalar to scale the data. To improve the root-squared (R^2) score of the data, LinearDiscriminantAnalysis (LDA) was implemented, after binning the data into categorial bins (low, medium, and high) and handling missing values by imputing with the mean. Hyperparameter tuning was additionally explored to find the optimal combination of hyperparameters that yield the best performance. The resultant best R^2 score was 0.1962371674882475, which gives the accuracy with which the model can predict the injection rate delta, whilst the root mean squared error was 3.5829616838461003, which indicates, on average, how much the predicted values differ from the actual ones. This was a novel assessment of using ML techniques for this purpose, and to further improve the R^2 score feature engineering can be explored by adding interaction terms, domain-specific or polynomial features, which can map out underlying patterns more effectively and thus improve the predictive capability of the program.

5. Research Directions

The field of CCS, incorporating ML techniques, holds great potential, with operational and cost-related challenges slowly diminishing. As the technology evolves to become more efficient, reliable and scalable, several promising research directions lay ahead.

5.1 Enhancing Data Quality and Availability

To train the various models, it is vital to ensure the availability of high-quality, comprehensive datasets. This should be a focal point of future CCS projects, which should build datasets incorporating geological data, operational parameters, and environmental conditions to accurately train and validate the ML models. This can be facilitated by the collaboration between academia, government, and industrial agencies. All of this will help improve the generalisability and accuracy of ML algorithms.

5.2 Advanced ML Algorithms for Real-Time Monitoring

Real-time monitoring of CCS operations is vital to ensure the safety and efficiency of the process. Research should explore the development of advanced ML algorithms capable of processing real-time data from various sensors and monitoring devices, using techniques such as RF for real-time monitoring and anomaly detection, SVM for fault detection and classification, and ANN for predictive maintenance.

5.3 ML-driven Optimisation of CO_2 capture techniques

Current CO_2 capture technologies, including absorption, adsorption, and membrane separation, can benefit from ML-driven optimisation. Research should focus on applying ML to optimise the design and operation of these technologies, which would improve their efficiency and thus reduce energy consumption. An example of this would be using ML models to predict the optimal operating conditions for chemical solvents used in absorption processes, or identifying the most effective materials for adsorption and membrane technologies.

5.4 Improving Storage Site Characterisation and Selection

The accurate characterisation and subsequent selection of CO_2 storage sites are critical for ensuring the long-term security of the stored CO_2. Future research should leverage ML techniques for analysing geological data such as porosity, permeability, and structural integrity of geological formations to predict the suitability of potential storage sites, using techniques such as ANN. Examples have already been outlined on how ML such as an ANN-GCS can be used to aid in developing predictive models to assess the long-term behaviour of CO_2 in storage sites, being able to identify potential leakage pathways over periods of nearly 300 years.

5.5 Enhancing Predictive Maintenance and Leak Detection

The transportation and storage phases of CCS are susceptible to leaks and equipment failures, which hold the risk of undermining the effectiveness of the entire process. Research should focus on developing ML-based predictive maintenance systems that can forecast equipment failures and schedule timely maintenance interventions, using models such as GAN to generate synthetic data. Additionally, advanced ML algorithms can be employed to improve leak detection systems, utilising data from various sensors and monitoring devices to quickly identify and mitigate leaks in pipelines and storage sites.

5.6 Integration of ML with Other Emerging Technologies

The convergence of ML with other emerging technologies, such as the Internet of Things (IoT), blockchain, and edge computing, presents exciting research opportunities. IoT devices can provide continuous streams of data from CCS operations, which can be analysed by ML algorithms like SVM and RT in real time. Data security and integrity can be further enhanced by blockchain technology, while edge computing can facilitate the processing of data at the source, thereby reducing latency and improving the responsiveness of ML-driven systems.

ML techniques can be seamlessly integrated across all these stages, from analysis, capture, and storage to ensure cost-effective CCS, with IoT facilitating the linking of these complex systems. This reduces costs associated with inefficient capture or storage leakage, which makes CCS practical in the long term, especially as its demand increases.

5.7 Addressing Ethical and Regulatory Challenges

The application of ML in CCS also raises ethical and regulatory challenges, which need to be addressed as the field matures, to ensure public trust, as consumers become more environmentally conscious. Research should explore the development of ethical frameworks and regulatory guidelines for the development of ML in CCS, which should include ensuring data privacy, mitigating biases and 'hallucinations' in ML models, and establishing accountability for ML-driven decisions; all of this requires a collaborative effort between researchers, policymakers, and industry stakeholders.

5.8 Future Directions in Interdisciplinary Research

Interdisciplinary research combining expertise in geoscience, chemical engineering, computer science, and environmental policy is crucial for advancing ML applications in CCS. Collaborative projects that integrate knowledge from these diverse fields can lead to innovative, novel solutions, and accelerate the effective adoption of ML in CCS. Funding agencies and research institutions should prioritise interdisciplinary research initiatives and support collaborative efforts to address the complex challenges of CCS.

6. Conclusion

The application of ML in CCS is still evolving, with ongoing research aimed at improving data quality, model interpretability, and integration with existing CCS infrastructure. Future developments may focus on the use of more advanced DL models, real-time data analytics, and the incorporation of ML-driven automation in CCS operations (Boot-Handford et al., 2014)[6].

Furthermore, the integration of ML with other emerging technologies, such as the Internet of Things (IoT) and blockchain, could revolutionize CCS by enabling more precise monitoring, better data security, and more efficient resource management (Yao et al., 2023)[11].

In conclusion, ML techniques offer transformative potential for CCS, enhancing the efficiency, reliability, and scalability of carbon capture, transportation, and storage processes. Continued research and innovation in this interdisciplinary field are essential for realizing the full potential of ML in combating climate change through effective CCS implementation. The future of this technology requires the training of models with large datasets to increase metrics like the R^2 score, which would enhance their accuracy and utility, allowing for a cleaner future to be in reach.

7. Bibliography

[1] Intergovernmental Panel on Climate Change (IPCC). (2021). Climate Change 2021: The Physical Science Basis. *Cambridge University Press*.

[2] Adjiman, C., Bardow, A., Anthony, E., Boston, A., Brown, S., Fennell, P., . . . Bui, M. (2018). Carbon capture and storage (CCS): the way forward. *Energy & Environmental Science*, Royal Society of Chemistry.

[3] Shreyash, N., Sonker, M., Bajpai, S., Tiwari, S., Khan, M., Raj, S., . . . Biswas, S. (2021). The review of carbon capture-storage technologies and developing fuel cells for enhancing utilization. *Eenrgies*.

[4] Yao, P., Yu, Z., Zhang, Y., & Xu, T. (2023). Application of machine learning in carbon capture and storage: An in-depth insight from the perspective of geoscience. *Fuel*.

[5] Budinis, S., Krevor, S., Mac Dowell, N., Brandon, N., & Hawkes, A. (2018). An assessment of CCS costs, barriers and potential. *Energy Strategy Reviews*, 61-81.

[6] Boot-Han (Anderson & Newell, 2004)dford, M., Abandes, J., Anthony, E., Blunt, M., Brandani, S., Mac Dowell, N., . . . Hallett, J. (2014). Carbon capture and storage update. *Energy & Environmental Science*, 130-189.

[7] Rahimi, M., Moosavi, S. M., Smit, B., & Hatton, T. A. (2021). Towards smart carbon capture with machine learning. *Cell Reoirts Physical Science*.

[8] Pires, J., Martins, F., Alvim-Ferraz, M., & Simões, M. (2011). Recent developments on carbon capture and storage: An overview. *Chemical Engineering Research and Design*, 1446-1460.

[9] Anderson, S., & Newell, R. (2004). Prospects for carbon capture and storage technologies. *Annual Review of Environment and Resources*, 109-142.

[10] Gibbins, J., & Chalmers, H. (2008). Carbon capture and storage. *Energy Policy*, 4317-4322.

[11] Yan, Y., Borhani, T. N., Subraveti, S. G., Pai, K. N., Prasad, V., Rajendran, A., . . . Clough, P. T. (2021). Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS) – a state-of-the-art review. Royal Society of Chemistry.

[12] Vaziri, P., & Sedaee, B. (2023). A machine learning-based approach to the multiobjective optimization of CO2 injection and water production during CCS in a saline aquifer based on field data. *Energy Science & Engineering*.

[13] Song, Y., Sung, W., Jang, Y., & Jung, W. (2020). Application of an artificial neural network in predicting the effectiveness of trapping mechanisms on CO_2 seuqestration in saline aquifers. *International Journal of Greenhouse Gas Control*.

[14] Nassabeh, M., You, Z., Keshavarz, A., & Iglauer, S. (2024). *Sub-surface geospatial intelligence in carbon capture, utlization and storage: A machine learning approach for offshore storage site selection.* Joondalup: Centre for Sustainable Energy and Resources, School of Engineering, Edith Cowan University. [14] Yan, Y., Borhani, T.

N., Subraveti, S. G., Pai, K. N., Prasad, V., Rajendran, A., . . . Clough, P. T. (2021). Harnessing the power of machine learning for carbon capture, utilisation, and storage (CCUS) – a state-of-the-art review. *Royal Society of Chemistry*.

[15] Song, Y.-y., & Lu, Y. (2015, April 25). *Decision tree methods: applications for classification and prediction.* Retrieved from National Library of Medicine: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/

[16] Balabin, R. M., & Lomakina, E. I. (2010). Support vector machine regression (SVR/LS-SVM) – An alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectrosco…. *The Royal Society of Chemistry 2011*.

[17] Hengrui, Z., & Olegovna, G. Y. (n.d.). *K-Means Clustering Algorithm And Improvement Methods.* Retrieved from https://libeldoc.bsuir.by/bitstream/123456789/56868/1/Zhang_Hengrui_K-means.pdf

[18] Machine Learning Challenge – *Using AI to Validate Carbon Containment in the Illinois Basin*. (n.d.). Retrieved from Think Onward: https://thinkonward.com/app/c/challenges/using-ai-to-validate-carbon-containment-in-the-illinois-basin#Data%20Source

[19] Wiberforce, T., Baroutaji, A., Soudan, B., Al-Alami, A., & Olabi, A. (n.d.). Outlook of Carbon Capture Technology and Challenges.