CO_2/brine relative permeability estimation using effective rock/fluid properties: A machine learning-based approach

Fatemeh Reisi, *Omid* Mohammadzadeh, *Lesley A.* James*

Department of Process Engineering, Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada

Abstract. Relative permeability is a crucial parameter in CO_2 storage as it describes the fluids' transport in porous media and residual saturations. CO_2 injection into saline aquifers leads to CO_2 dissolution in brine and subsequent formation of carbonic acid, resulting in complicated reactions with the rock minerals. These reactions include mineral dissolution, mineral precipitation or silicate weathering, depending on the composition and texture of the reservoir rock. As a consequence, these reactions bring about alterations in key rock characteristics including mineralogy, porosity, absolute permeability, and wettability. The main objective of this study is to develop a CO_2/brine relative permeability model using machine learning (ML) algorithms based on a more comprehensive, yet representative, set of input parameters including rock/fluid properties and their compositions, instead of time-consuming experiments. In this study, experimental relative permeability data were data mined from the literature, along with some additional fluid and rock properties including absolute permeability, porosity, fluid density, fluid viscosity, CO_2/brine interfacial tension, CO_2 solubility, rock minerals' concentration, brine salinity and composition, and fluid saturations. The dataset was pre-processed, and missing data were identified and populated using applicable correlations. Various supervised ML approaches were tried to predict the relative permeability values, including Decision Tree Regression (DTR), Gaussian Process Regression (GPR), Linear Regression (LR), and Bayesian Ridge Regression (BRR). In a comparative analysis, the same ML algorithms were trained and tested, incorporating only three typical input parameters that have been repeatedly used recently in the literature for relative permeability modelling (i.e., porosity, absolute permeability, and fluids saturation). Results indicate that there was a good agreement between the predicted and experimental relative permeability data for models developed using DTR and GPR algorithms when all input parameters were incorporated. However, the models were not nearly as accurate when only three conventionally agreed input parameters were used. A sensitivity analysis was also performed to determine the most accurate input parameters and their impact on relative permeability predictions. According to the sensitivity analysis results, the most important input parameters are brine salinity, CO_2 and water viscosity, porosity, absolute permeability, and iron and halite content of the rock samples. Overall, the results indicate the critical role of rock-fluid interactions and composition, as well as fluid properties, in predicting relative permeability using ML-based approaches, underscoring their significance for future studies.

1 Introduction

Carbon capture, utilization, and storage technologies are efficient solutions to reduce environmental pollution, climate change, and global warming. CO_2, as one of the most important greenhouse gases, can be stored in oceans, geological structures, or through surface mineral carbonation. Since surface mineral carbonation and CO_2 storage in oceans require expensive infrastructure and may even negatively affect the environment, CO_2 storage in geological formations seems to be the best option. Moreover, potential geological storage sites are distributed worldwide, and all the experience and knowledge obtained from fossil fuel production and geological studies are applicable to geological CO_2 storage. Potential geological sites for CO_2 storage include depleted oil and gas reservoirs, deep saline aquifers, and coal seams. Saline aquifers are of utmost importance when geological sites for carbon storage are screened due to their large capacity, wide geographical spread, and reasonable porosity and absolute permeability [1, 2].

Understanding how different fluids flow through a rock can be achieved by examining the multiphase flow properties of a porous medium-fluid(s) system, including relative permeability and capillary pressure. These properties explain how each fluid phase is displaced and how different fluid phases are distributed within the pores. The ability to store CO_2 in both mobile and immobile forms within the pore space and in the presence of brine relies on the relative permeability of CO_2-brine as well as capillary pressure characteristics [3]. Relative permeability plays a crucial role in describing these parameters. Relative permeability is an important characteristic that influences the movement and fate of injected CO_2 in porous media. It describes how much the injected CO_2 and water disrupt each other's movement through rocks. Moreover, relative permeability curves are essential inputs for reservoir simulators, enabling them to history match and simulate CO_2 storage scenarios [4].

* Corresponding author: ljames@mun.ca

Misinterpretation of the CO_2-brine relative permeability could be the reason behind various unexpected field-scale observations, such as CO_2 injectivity issues, early breakthrough, and plume shape [5]. In the last few decades, the research focus on the geological storage of CO_2 has led to examining the CO_2/brine system in aquifer rocks, particularly in terms of relative permeability [6].

During CO_2 injection into aquifers, CO_2 dissolves in brine and reacts with the rock minerals. Carbonic acid is formed upon CO_2 dissolution in the brine, leading to reduction in pH and an increase in the H+ concentration. Thus, different reactions occur depending on the brine and rock compositions as well as reservoir conditions. For instance, mineral dissolution occurs when carbonic acid reacts with calcium carbonate minerals. Mineral precipitation could simultaneously happen when calcium and bicarbonate ions react, forming solid calcium carbonate. Silicate weathering is yet another reaction that occurs when carbonic acid reacts with silicate minerals. These reactions could cause changes in crucial rock and rock-fluid properties such as mineralogy, porosity, absolute permeability, and wettability through mineral dissolution/precipitation and surface chemistry alteration [7, 8]. For instance, specific mineral dissolution increases porosity and absolute permeability, while mineral precipitation and particle migration have the opposite effect. The concentration of ions in brine, type of rock minerals, and CO_2 injection pressure, as well as reservoir pressure, all play significant roles in determining the extent of these reactions [9].

The above-mentioned reactions and subsequent changes in rock and rock-fluid properties have implications for relative permeability, which has been an active area of research for the past several years; however, there is a lack of site-specific reservoir data for CO_2 storage application. Particularly, several papers have been published over the years on determining CO_2/brine relative permeability through laboratory experimentation [10]. In these experiments, both unsteady-state and steady-state coreflooding techniques could be used to determine initial brine and residual CO_2 saturations, relative permeabilities during drainage and imbibition, and other system characteristics under various flow conditions, temperature, and pressure [1, 11]. Based on these experimental studies, different models (i.e., correlations) have been developed for relative permeability prediction; however, specific impacts of fluid composition, properties as well as rock-fluid interactions on relative permeability curves or saturation exponents have not been fully understood and implemented yet in these so-called conventional models for relative permeability prediction. These conventional models mainly describe the relative permeability as a function of fluid phase saturations. There are more recent studies that consider some additional input parameters such as wettability, rock type, porosity, absolute permeability, viscosity ratio, and interfacial tension [6, 12, 2].

In this study, we developed some machine learning (ML)-based models for predicting CO_2/brine relative permeability by utilizing a comprehensive set of input parameters that incorporate rock and fluid properties and compositions, as well as rock-fluid interactions. The input parameters used in this ML-based modelling attempt include absolute permeability, porosity, fluid density, fluid viscosity, CO_2/brine interfacial tension, CO_2 solubility, rock mineral concentration, water salinity and composition, and fluid saturations. To the best of our knowledge, no such modelling exercise has been done in the literature for predicting the relative permeability of CO_2/brine fluid system that includes the comprehensive range of input parameters utilized in this research work. As pointed out above, the currently available predictive models only consider a limited range of input parameters, which results in reduced accuracy and non-representativeness of the models.

2 Methodology

A comprehensive data gathering and databank preparation was done in this research work in which several papers focusing on CO_2-brine relative permeability measurements [1-30] were reviewed, and all the data, including rock and fluid properties, rock-fluid interaction parameters as well as relative permeability information were extracted accordingly. The parameters extracted include absolute permeability, porosity, fluid density, viscosity, CO_2/brine interfacial tension, CO_2 solubility, rock mineral concentration, water salinity and composition, fluid saturations, and corresponding relative permeability values. The data collected were pre-processed, and missing data such as fluid density and viscosity, and CO_2/brine interfacial tension were identified and populated using applicable correlations. Several supervised machine learning approaches were attempted to predict the relative permeability values at corresponding fluid saturations, including Decision Tree Regression (DTR), Gaussian Process Regression (GPR), Linear Regression (LR), and Bayesian Ridge Regression (BRR). A subset of the collected data (25% of the dataset) was utilized as the training dataset to train the model. The accuracy and reliability of the models were then assessed using the remaining data, designated as the testing dataset. The data points allocated for the testing stage were excluded from the databank when the models were trained. Therefore, the accuracy of predictions achieved in the testing stage was not biased with any data shared between the training and testing stages. The accuracy of predictions with respect to the measured values was assessed by calculating some statistical parameters such as root-mean-square error (RMSE), relative average deviation (RAD), and coefficient of determination (R^2) based on model predictions during the testing stage. Additionally, the predicted versus measured relative permeability profiles were compared using some selected ML approaches as a function of aqueous phase saturation. A sensitivity analysis was also performed to determine the most influential input parameters and their impact on relative permeability predictions. In this study, the relative permeability predictions were obtained by incorporating all the input parameters into the ML approaches. The relative permeability values were also calculated with only three input parameters of porosity, absolute permeability, and fluid saturation.

2.1 Data pre-processing

As previously stated, a dataset consisting of nearly 700 data points was created, and the missing values, accounting for approximately 10% of the entire dataset, were identified using different correlations and filled into the database. Table

1 displays the missing parameters and the correlations used for estimating these missing parameters.

Table 1. Correlations used for estimating the missing parameters in order to complete the database

Missing Parameters in the database	Correlation	Correlation parameters
Brine Density	Danesh [31]	P, T, S
Brine Viscosity	Danesh [31]	P, T, S
CO_2 Viscosity	Pitzer & Schreiber [32]	P, T
CO_2 Density	Altunin & Sakhabetdinov [33]	P, T
CO_2 Solubility	Hangx [34]	P, T, concentration of salts
IFT	Lins et al. [35]	P, T, ρ_{CO2}, ρ_{brine}

P: Pressure
T: Temperature
S: Brine salinity
ρ: Density

After filling in the missing values, the descriptive statistical distributions for each input parameter were computed. The range of variation for each input parameter is illustrated in Fig. 1 in the logarithm scale. Clearly, certain parameters, such as salinity and absolute permeability exhibited a wider variation range compared to the rest of the input parameters.

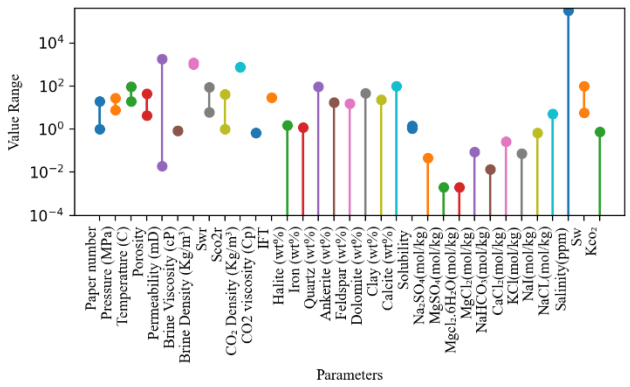

Fig. 1. Variation range of input parameters

A descriptive statistical analysis approach was also undertaken to gain deeper insights into the dataset. This involved computation of various statistical parameters such as mean, standard deviation, minimum, maximum, and quartiles for all the input parameters. This approach enabled us to determine whether the values fell within a reasonable range and also offered insights into the overall distribution patterns. The calculated statistical parameters for some of the features are presented Table 2.

Table 2. Statistical parameters for some of the features

Statistical parameters \ Input	φ	k (mD)	ρ_{CO2} (kg/m³)	μ_{CO2} (cp)	S (ppm)
Mean	17.45	480.88	616.53	0.10	71,497
std	7.6	697.32	201.99	0.17	93,501
Min	4.3	0.02	209.9	0.02	0
25%	12.57	2.8	404	0.03	0
50%	16.8	65	680	0.05	30,000
75%	22.2	619	768.4	0.08	88,972
Max	44	1810	920	0.66	315,400

k: Absolute permeability
φ: Porosity
μ: Viscosity

2.2. Data preparation for modelling

An essential step in ML modelling is dividing the input data into training and testing subsets. Cross-validation is a technique employed in ML modelling where the training and testing subsets are used alternately as both a testing and validation set, with the process repeated multiple times to assess the model's performance. The objective is to evaluate the model's ability to generate accurate and reliable predictions based on measured values. The cross-validation procedure provides a more precise estimation of the model's performance by utilizing multiple validation datasets. Another approach for creating data subsets is a straightforward strategy in which the entire dataset is randomly divided into separate training and testing subsets. In this research work, we employed both approaches. We first incorporated a 5-fold cross-validation method. Additionally, 25% of the data were randomly allocated for model training, and then the remaining 75% were used as the testing subset.

2.3. ML Algorithms

Once the training and testing datasets were prepared, ML algorithms were trained and evaluated. The subsequent four sections present a concise overview of each of the employed ML applications.

2.3.1 Linear Regression

Linear regression is a widely-used and straightforward method that relies on mathematical approaches to perform predictive analysis. It is particularly useful for projecting continuous or mathematical variables. LR is a statistical method to model the relationship between a dependent variable and one or more independent variables. The goal of using LR is to find the best-fitting linear relationship between the dependent and independent variables such that the difference between the predicted and actual values of the dependent variable is minimized. LR can be classified into two categories: simple linear regression and multiple linear regression (MLR). A straight line is used in simple linear regression to model the relationship between the dependent variable and ONE independent variable. However, in MLR, a hyperplane is used to model the relationship between the dependent variable and two or more independent variables

[36]. The model's parameters are estimated using the least squares method in both cases. An intercept and a slope for simple linear regression define the model. In contrast, a MLR model is defined by a set of parameters, one for each independent variable. This method is easy to implement, and it can also be extended to more complex models, such as polynomial regression or generalized linear models. However, it assumes a linear relationship between the variables and may not perform well when the relationship is non-linear, or the data contain outliers [37].

2.3.2 Decision Tree Regression

One of the practical approaches to supervised learning is Decision Tree. This algorithm applies to both regression and classification problems. The decision tree algorithm is a decision-making tree-like structure technique that divides a dataset into increasingly smaller subsets while simultaneously developing a corresponding decision tree. The resulting tree consists of decision nodes and leaf nodes. The first node is the root node, representing the entire sample, and may be divided into further nodes [38]. The interior nodes are associated with the features of a dataset, and the branches represent decision rules. The leaf nodes are the outcomes. DTR is utilized where the goal is to predict a continuous numerical value. Once the tree is constructed, a new data point is predicted by traversing the tree from the root to a leaf node. This algorithm evaluates the condition based on the input feature at each internal node and decides whether to follow the left or right branch. The leaf node reached by the traversal then represents the predicted value for the input data point [39].

2.3.3 Gaussian Process Regression (GPR)

The Gaussian Process is a versatile technique for supervised learning that can be applied to solve regression and probabilistic classification problems. GPR is a regression technique that is nonparametric and nonlinear. It is utilized to interpolate between data points dispersed in high-dimensional input space. The algorithm is rooted in Bayesian probability theory and strongly correlates with other regression approaches, such as kernel ridge regression (KRR) and LR. In contrast to several widely used supervised machine learning algorithms which learn precise values for each parameter in a function, the Bayesian methodologies instead estimate a probability distribution encompassing all the potential values. It is particularly useful when little or no prior knowledge about the underlying function exists. GPR provides not only point estimates but also incorporates uncertainty estimates for the predictions, making it suitable for a range of applications. Additionally, GPR is applicable to handling data with complex input-output relationships and can be employed for tasks such as interpolation, extrapolation, and function approximation [40].

2.3.4 Bayesian Ridge Regression (BRR)

Bayesian Regression is a statistical methodology that utilizes probability distributions instead of point estimates to model LR. This technique is particularly effective in handling data that may be inadequate or non-uniformly distributed. Unlike conventional methods that estimate the output variable as a single value, Bayesian Regression models the output variable as a probability distribution. Thus, the algorithm increases the modeling flexibility and considers data uncertainty. BRR is a widely used technique among Bayesian linear regression algorithms that includes a regularization term to improve the model's performance. Essentially, BRR reduces the impact of noise and extraneous variables by placing a prior probability distribution on the regression coefficients; therefore prevents overfitting and improves the overall model predictability. In particular, BRR is highly effective in dealing with many independent variables when some are interrelated [41].

3 Result and Discussion

Four supervised ML algorithms were used to model CO_2/brine relative permeability as a function of absolute permeability, porosity, fluid density, fluid viscosity, CO_2/brine interfacial tension, CO_2 solubility, rock mineral concentration, water salinity and composition, and fluid saturations. Data pre-processing and understanding/finding missing data are crucial steps for proper data analysis. A Python-generated heatmap chart is introduced in Fig. 2 to visualize the extent of available data and facilitate data pre-processing. In Fig. 2a, each cell represents a data point, and the missing values are depicted in white while the available data are illustrated in black. Significant clusters of missing data can be observed for a few input parameters including CO_2, water, and total flow rates, and CO_2 solubility. In addition, certain data points were also missing in some other variables, such as interfacial tension (IFT), fluid viscosities, and densities. It was not possible to correlate/estimate the missing unreported injection flow rates of CO_2 and water based on available information from the reference papers. Therefore, it was decided to exclude these parameters from further analysis. For the remaining missing parameters, however, we managed to estimate them using some appropriate correlations to ensure a comprehensive dataset is available for further analysis. After excluding the injection flow rates and incorporating missing data estimation using proper correlations, the resulting dataset contains no missing information, as depicted in Fig. 2b.

After cleaning the dataset and filling in the missing values, all the input data were normalized to ensure that features with larger magnitudes do not dominate or bias the learning algorithm. As depicted in Fig. 1 and 3, brine salinity exhibits notably larger values compared to other features when the original data points are concerned. With data normalization, all the data points were brought within a consistent range of variation in the input dataset.

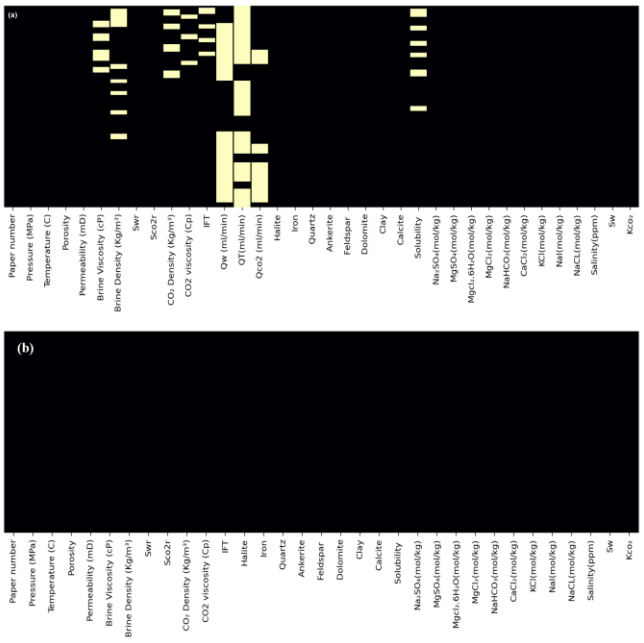

Fig. 2. Heatmap showing the presence of null values in the dataset. The white colour represents the missing data.

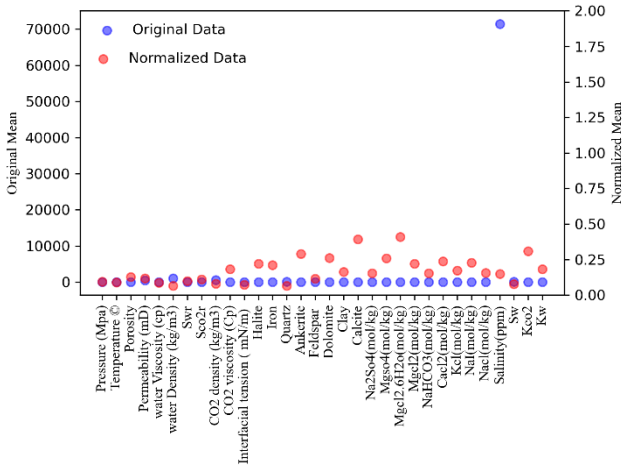

Fig. 3. Comparison of "feature mean" for the original and normalized datasets, showing the impact of normalization on data distribution.

After data pre-processing and normalization, the ML algorithms were first trained and then individually used to predict CO_2 relative permeability for the whole dataset, and the agreements between the predicted and measured relative permeability values for each ML model are presented in the form of parity plots in Fig. 4 to 7. In each parity plot, the solid black line represents the line where the measured and predicted data are equal. The two black dashed lines represent the predicted values within the ±5% range from the measurements. Clearly, the DTR and GPR algorithms demonstrated significant superiority in predicting the measured data when compared to the other methods. As indicated in Fig. 4 and 5, the scattering errors associated with these predictions are small and very similar, and most data points are spread closely around the diagonal line. In contrast, the LR and BRR models possess less accuracy when predicting CO_2 relative permeability values. This is evident from the wider distribution of the data points beyond the ±5% scatter error zone depicted by the dotted lines in Fig. 6 and 7.

In order to quantitatively assess the predictive performance of each of these ML algorithms, a set of statistical parameters was calculated, including root-mean-square error, relative average deviation, and coefficient of determination (Table 3). These statistical measures show the accuracy of predictions when compared against experimental data. Clearly, the DTR and GPR algorithms resulted in the smallest RMSE and RAD values and the biggest R^2 values, which signify their highest accuracy in predicting CO_2 relative permeability values. Interestingly, both the parity plots and statistical measures show that the DTR model slightly outperformed the GPR model in predicting CO_2 relative permeability values.

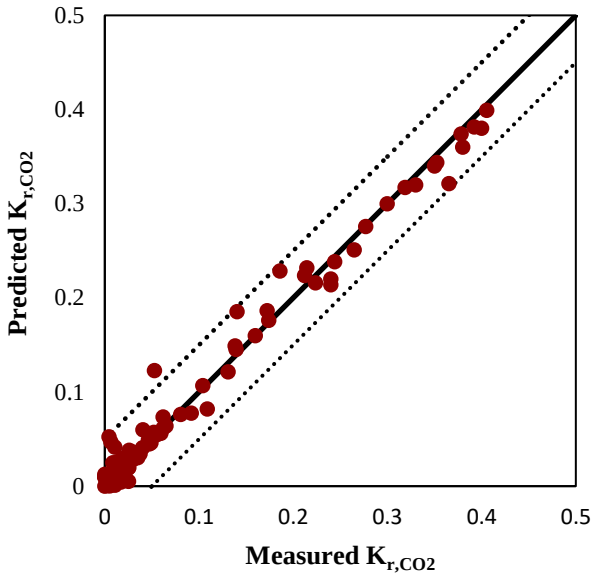

Fig. 4. Comparison of the predicted and measured CO_2 relative permeability values using DTR model.

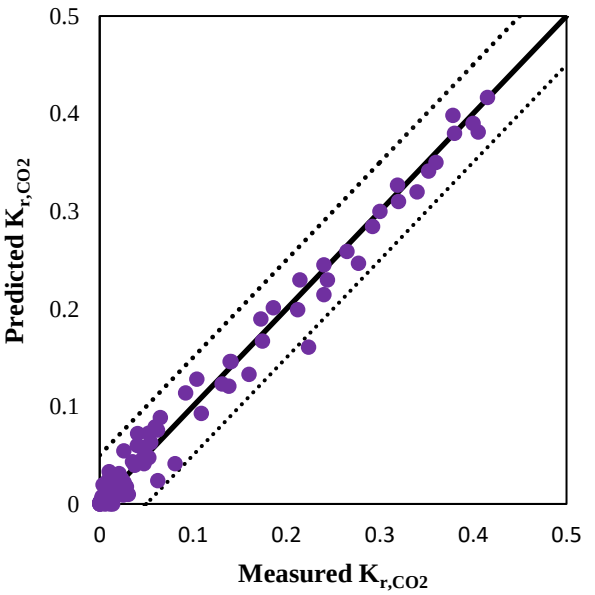

Fig. 5. Comparison of the predicted and measured CO_2 relative permeability values using GPR model.

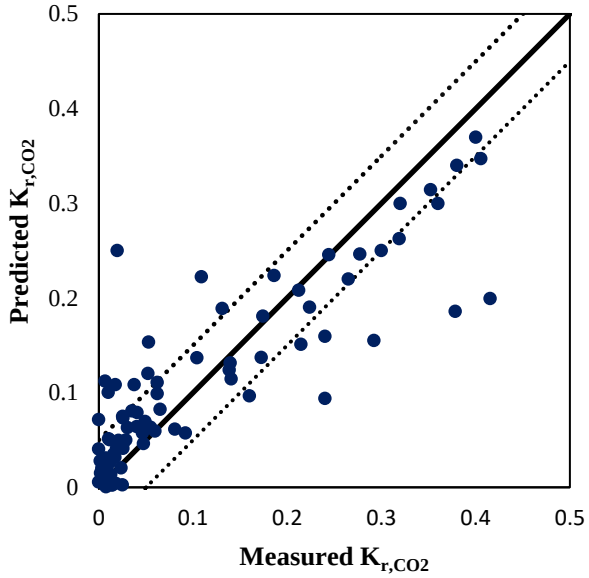

Fig. 6. Comparison of the predicted and measured CO_2 relative permeability values using LR model.

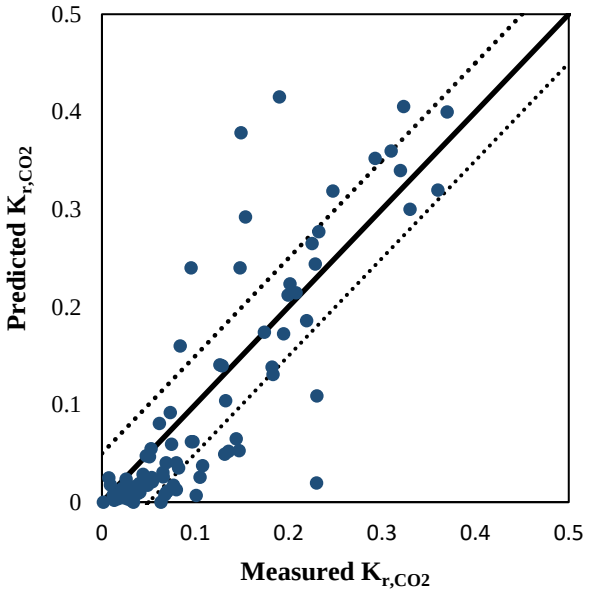

Fig. 7. Comparison of the predicted and measured CO_2 relative permeability values using BRR model.

Table 3. Statistical measures presenting models' accuracy.

Method / Algorithm	RMSE	RAD (%)	R^2
BRR	0.086	0.49	0.62
LR	0.08	0.48	0.64
DTR	0.01	0.0212	0.96
GPR	0.025	0.0241	0.94

Another approach to demonstrate the excellent performance of the DTR and GPR models in predicting the CO_2 relative permeability data is through visualization of the residual error distribution. As demonstrated in Fig. 8, the residual errors associated with both CO_2 relative permeability predictions are normally distributed around zero, which shows high-accuracy predictions for both algorithms.

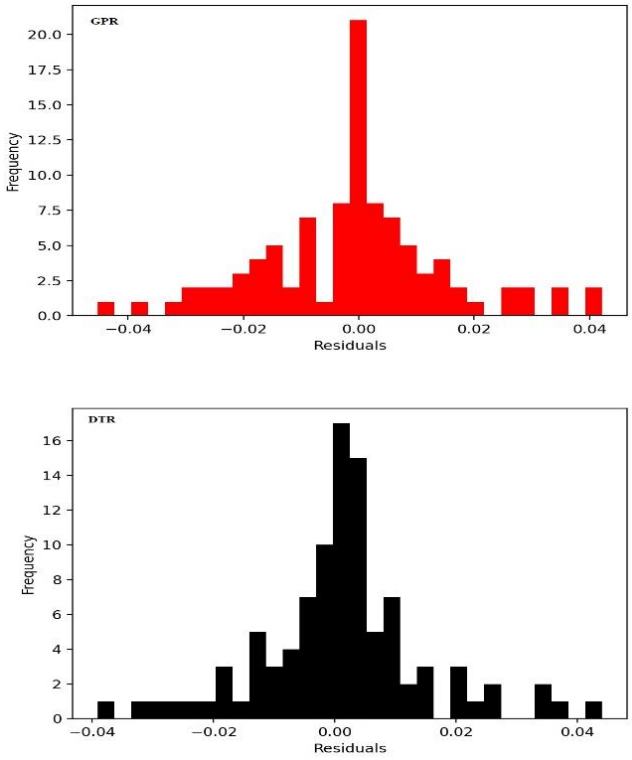

Fig. 8. Residual error distribution plots for the DTR and GPR algorithms.

Once the performance of the ML models was assessed in predicting the CO_2 relative permeability values for the entire input parameter dataset, the two most promising models of DTR and GPR were utilized to build the entire relative permeability-saturation function for both CO_2 and brine phases based on the data collected from three selected case studies in references [1], [4] and [9]. Note that the R^2 values reported in Fig. 9 to 11 belong to the CO_2 relative permeability predictions using the DTR and GPR models when compared against the experimental data. As indicated in Fig. 9 to 11, the selected ML models were able to replicate the experimental data very accurately, with coefficients of determination of greater than 0.96, over the entire range of CO_2 saturation, with only a slight underestimation at the higher end of CO_2 saturations. The collective prediction trends shown for the DTR and GPR models strongly suggest them as two representative models for predicting CO_2 relative permeability for the CO_2-brine fluid system. As indicated in these three figures, the DTR and GPR models also predicted the brine relative permeability very accurately. In order to emphasize the outstanding performance of the DTR and GPR models, one set of CO_2 relative permeability values versus saturation, borrowed from Reynolds et al. [4], was compared against predictions made using the LR and BRR models (Fig. 12). It is clear that these two models did not perform well in predicting CO_2 relative permeability values.

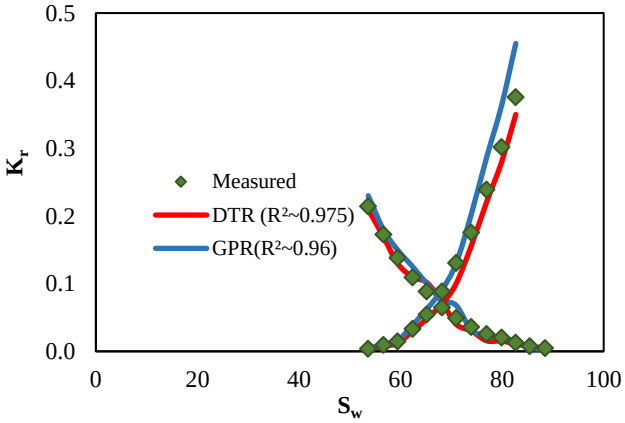

Fig. 9. Comparison of measured (Kou et al. [1]) versus predicted CO_2 and brine relative permeability curves with the application of DTR and GPR models.

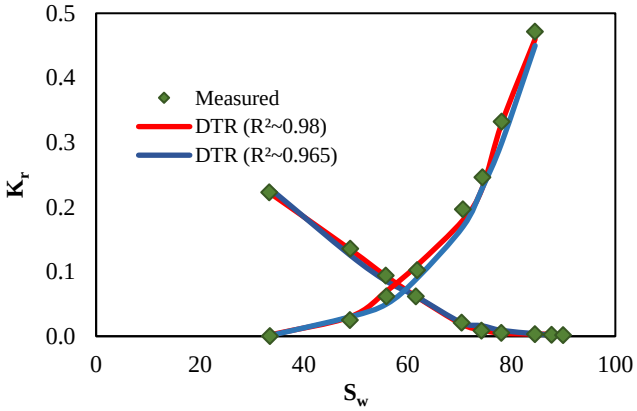

Fig. 10. Comparison of measured (Reynolds et al. [4]) versus predicted CO_2 and brine relative permeability curves with the application of DTR and GPR models.

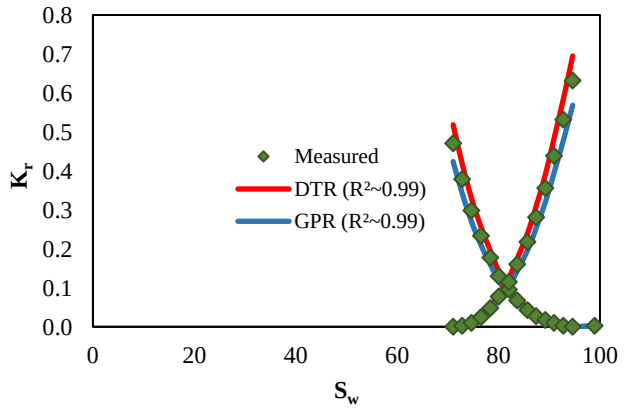

Fig. 11. Comparison of measured (Kou et al. [9]) versus predicted CO_2 and brine relative permeability curves with the application of DTR and GPR models.

In recent years, various researchers used three parameters of saturation, absolute permeability and porosity to predict relative permeability through analytical and numerical modelling efforts [6, 12, 42, 43]. We propose that the prediction accuracy will be significantly enhanced if more carefully-selected relevant input parameters are incorporated when modelling relative permeability using ML approaches. To prove this hypothesis, the Decision Tree and Gaussian Process Regression models were trained with only three parameters of porosity, absolute permeability and fluid saturation. These models were then used to predict the CO_2 relative permeability values, and the predicted versus measured relative permeability results are plotted in Fig. 13 and 14. In each parity plot, the solid black line represents the line where the measured and predicted data are equal, and the two black dashed lines represent the predicted values within the ±5% range from the measurements. Clearly, the prediction performance of both these ML algorithms has deteriorated when compared to their prediction accuracy with all the input parameters involved for training and testing of the models.

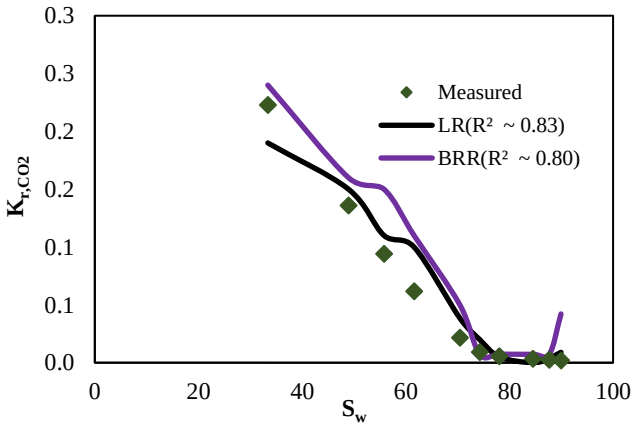

Fig. 12. Comparison of measured (Reynolds et al. [4]) versus predicted CO_2 relative permeability curves with the application of LR and BRR models.

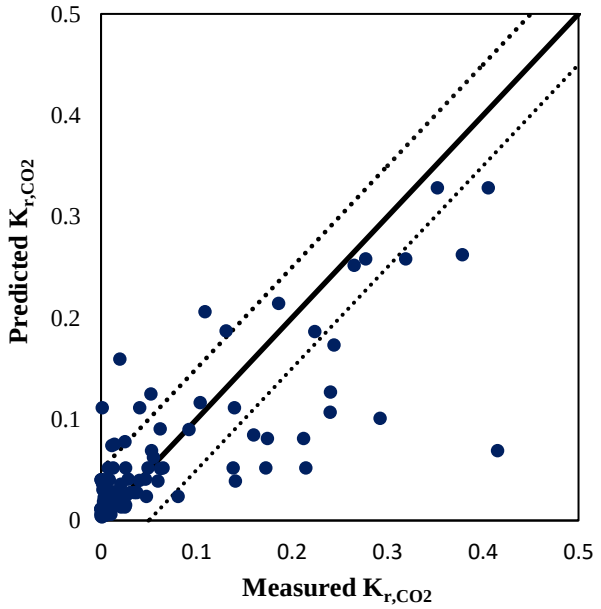

Fig. 13. Comparison between the predicted and measured CO_2 relative permeability values with application of DTR algorithm when only three parameters of porosity, absolute permeability and fluid saturation were used for the modelling approach.

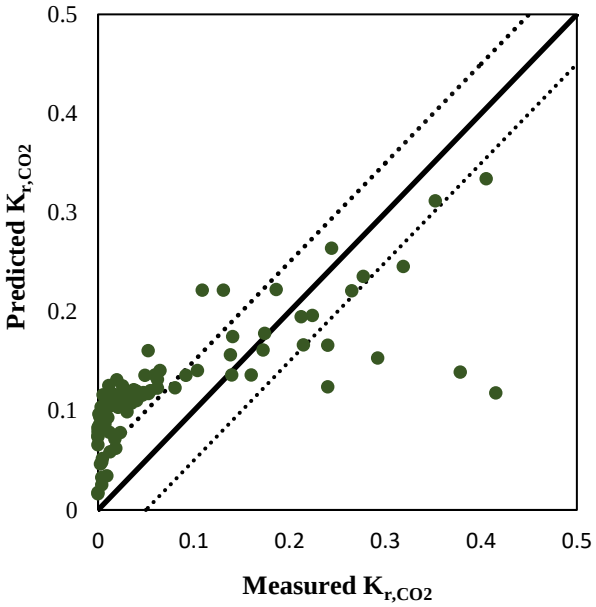

Fig. 14. Comparison between the predicted and measured CO_2 relative permeability values with application of GPR algorithm when only three input parameters of porosity, absolute permeability and fluid saturations were used for the modelling approach.

In order to assess the relative significance of various input parameters on the predictive performance of each model, a sensitivity analysis was conducted using the DTR and GPR algorithms, and the results are presented in Fig. 15 and 16, respectively. The most influential parameters obtained from this sensitivity analysis include brine saturation and salinity, CO_2 and water viscosity, porosity, absolute permeability, and iron and halite concentration of rock. These results highlight the importance of considering these influential parameters in future research and the practical application of the model. The subsequent paragraphs briefly discuss these parameters' significance with some evidence from the literature.

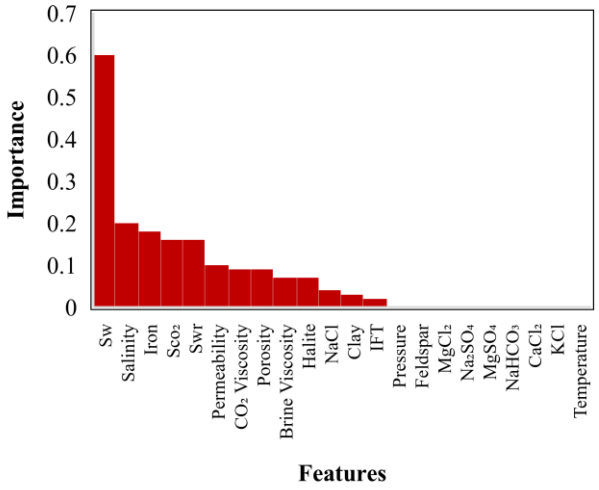

Fig. 15. Sensitivity analysis of the relative permeability results using the DTR algorithm.

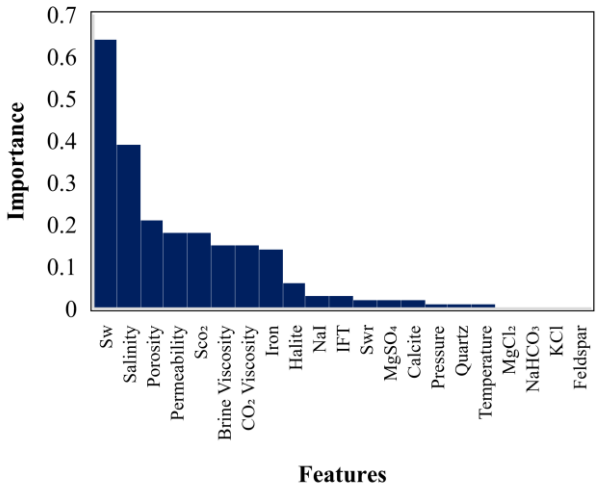

Fig. 16. Sensitivity analysis of relative permeability results using the GPR algorithm.

Salinity is known to influence rock wettability and, therefore, is an important factor when the fluid distribution and relative permeability measurements are concerned. It also inversely affects CO_2 solubility in brine. Various chemical reactions, such as the ion exchange phenomenon at the fluid-rock interface, are highly affected by brine salinity. It is customary to have salinity as one of the most influential parameters for relative permeability prediction [44, 45]. Rock porosity and absolute permeability are two other important parameters recently considered to be "influential" when predicting relative permeability. The interconnected pore network within a rock structure significantly affects porous media flow. Porosity, along with pore size distribution, determines pore connectivity, hence affecting rock permeability. Additionally, porosity demonstrates the effective surface area available for fluid-rock interactions. Greater porosity values provide a larger contact area between CO_2, brine, and the rock surfaces, potentially leading to complex interactions affecting relative permeability [44].

The presence of various minerals (such as iron) in rocks could facilitate mineral dissolution and subsequent re-precipitation (due to pH changes) when CO_2 reacts with brine and reservoir rock, leading to potential changes in relative permeability. The presence of certain minerals could also affect surface wettability when in contact with the CO_2-brine fluid system [46]. The next influential parameter is the halite content of the reservoir rock. Halite, also known as rock salt, mainly comprises sodium chloride (NaCl). The presence of halite in reservoir rocks significantly affects CO_2 and brine relative permeabilities due to its high solubility in water [47]. Last but not least, CO_2 and brine viscosities significantly affect flow resistance, fluid mobility, saturation distribution, and capillary pressure gradients within the porous medium. These factors collectively affect the relative permeabilities associated with the CO_2-brine fluid system [37].

4 Conclusions

This study aimed to predict CO_2/brine relative permeability using supervised machine learning algorithms. Four

algorithms were employed: Linear Regression, Decision Tree Regression, Gaussian Processes Regression, and Bayesian Ridge Regression. The Decision Tree and Gaussian Process Regression algorithms accurately modelled the relative permeability dataset (R^2 > 94%). The performance of these algorithms was further validated by predicting relative permeability pairs versus saturation profiles for three case studies borrowed from the literature. Additionally, a sensitivity analysis was performed to investigate the effect of selecting only a limited number of input parameters on the predicted relative permeability data. It was concluded that more accurate relative permeability predictions were obtained when the full list of input parameters was used. Moreover, a sensitivity analysis was performed to obtain the most influential parameters when predicting relative permeability data using a selected ML algorithm. This list includes brine salinity, CO_2 and water viscosity, porosity, absolute permeability, as well as iron and halite concentration of the reservoir rock. In summary, this study highlights the ability of some ML algorithms to accurately predict the relative permeability data using some rock, fluid, and rock/fluid parameters. It also highlights the importance of considering a comprehensive set of input parameters to provide thorough insights into relative permeability prediction with high certainty levels.

Acknowledgements

The authors would like to thank the Hibernia Management and Development Company (HMDC), Chevron Canada Ltd, Energy Research and Innovation Newfoundland and Labrador (ERINL), the Natural Sciences and Engineering Research Council of Canada (NSERC), the Province of Newfoundland and Labrador, and Mitacs for financial support.

References

1. Z. Kou, H. Wang, V. Alvarado, J. F. McLaughlin, S. A. Quillinan, *J. Hydrol*. **599**, 126481 (2021)
2. G. S. Jeong, J. Lee, S. Ki, D. G. Huh, C. H. Park, *Energy*, **133,** 62-69 (2017)
3. R. Farokhpoor, E. G. B. Lindeberg, O. Torsæter, M. B. Mørk, A. Mørk, *Sci. Technol*. **4(1)**, 36-52 (2014)
4. C. Reynolds, M. Blunt, S. Krevor, *Energy Procedia*, ***63***, 5577-5585 (2014)
5. C. A. Reynolds, S. Krevor, *Water Resour. Res.* **51(12)**, 9464-9489 (2015)
6. J. C. Manceau, J. Ma, R. Li, P. Audigane, P. X. Jiang, R. N. Xu, C. Lerouge, *Water Resour. Res.* **51(4)**, 2885-2900 (2015)
7. Y. Tang, S. Hu, Y. He, Y. Wang, X. Wan, S. Cui, K. Long, *J. Chem. Eng.,* **413**, 127567 (2021)
8. R. J. Rosenbauer, T. Koksalan, J. L. Palandri, *FPT*, **86**, 1581-1597 (2005)
9. Z. Kou, H. Wang, V. Alvarado, C. Nye, D. A. Bagdonas, J. F. McLaughlin, S. A. Quillinan, *SPE*, **28(02**), 754-767 (2023)
10. S. A. Smith, C. J. Beddoe, B. A. Mibeck, L. V. Heebink, B. A. Kurz, W. D. Peck, L. Jin, *Energy Procedia*, **114**, 2957-2971 (2017).
11. M. Akbarabadi, M. Piri, *Adv Water Resour.* **52**, 190-206 (2013)
12. M. A. Ahmadi, *Fuel*, **140**, 429-439 (2015)
13. L. Jin, L. J. Pekot, S. A. Smith, O. Salako, K. J. Peterson, N. W. Bosshart, C. D. Gorecki, *Int. J. Greenh. Gas Control.*, **75**, 140-150 (2018)
14. P. Tawiah, H. Wang, S. L. Bryant, M. Dong, S. Larter, J. Duer, *Int. J. Greenh. Gas Control.*, **112**, 103491 (2021)
15. X. Zhang, Doctoral dissertation, *UNSW Sydney*, (2021)
16. M. Seyyedi, A. Giwelli, C. White, L. Esteban, M. Verrall, B. Clennell, *Fuel*, **269**, 117421 (2020)
17. S. A. Smith, C. J. Beddoe, B. A. Mibeck, L. V. Heebink, B. A. Kurz, W. D. Peck, L. Jin, *Energy Procedia*, **114**, 2957-2971 (2017)
18. S. Bachu, B. Bennion, *Environ. Geol.,* 54, 1707-1722 (2008)
19. L. Rasmussen, T. Fan, A. Rinehart, A. Luhmann, W. Ampomah, T. Dewers, R. Grigg, *Energies*, **12(19)**, 3663 (2019)
20. N. Liu, S. Ghorpade, L. Harris, L. Li, R. Grigg, R. Lee, *OnePetro* (2010)
21. S D. B. Bennion, S. Bachu, *SPE Reserv. Evaluation Eng.* **11(03)**, 487-496 (2008).
22. A. Saeedi, C. Delle Piane, L. Esteban, Q. Xie, *Int. J. Greenh. Gas Control.,* **54**, 309-321 (2016)
23. A. Al-Menhali, B. Niu, S. Krevor, *Water Resour. Res.*, **51(10),** 7895-7914 (2015)
24. X. Chen, A. Kianinejad, D. A. DiCarlo, *SPE Symp. Improv. Oil Recovery*. *OnePetro,* (2014)
25. S. C. Krevor, R. Pini, L Zuo, S. M. Benson, *Water Resour. Res.*, **48(2)** (2012)
26. B. Bennion, S. Bachu, *SPE Annu. Tech. Conf. Exhib. OnePetro* (2005)
27. B. Bennion, S. Bachu, *SPE Annu. Tech. Conf. Exhib. OnePetro* (2010)
28. S. Berg, S. Oedai, H. Ott, *Int. J. Greenh. Gas Control.*, **12**, 478-492 (2013)
29. N. Wei, M. Gill, D. Crandall, D. McIntyre, Y. Wang, K. Bruner, G. Bromhal, *Greenh. Gas. Sci. Technol.,* **4**, 1–19 (2014)
30. J. C. Perrin, S. Benson, *Transp. Porous Media*., **82**, 93-109 (2010)
31. A. Danesh, PVT and phase behaviour of petroleum reservoir fluids. *Elsevier*, (1998)
32. K. S. Pitzer, D. R. Schreiber, *Fluid Phase Equilib*., **41(1-2)**, 1-17 (1988)
33. V. V. Altunin, M. A. Sakhabetdinov. *Teploenergetika*, **8**, 85-89 (1972)
34. S. Hangx, Subsurface mineralisation. Rate of CO_2 mineralisation and geomechanical effects on host and seal formations. A review of relevant reactions and reaction rate data. *First interim report*. (2005)

35. I. E. Lins, G. P. Santana, G. M. Costa, S. A. V. de Melo, *Fluid Phase Equilib*., **555**, 113354 (2022)
36. D. Maulud, A. M. Abdulazeez, *JASTT*, **1(4)**, 140-147 (2020)
37. X. Su, X. Yan, C. L Tsai, *Wiley Interdiscip. Rev. Comput. Stat.*, **4(3)**, 275-294 (2012)
38. E. Frank, Y. Wang, S. Inglis, G. HolmesI. H. Witten, *Mach. Learn.*, **32**, 63-76 (1998).
39. B. Gupta, A. Rawat, A. Jain, A. Arora, N. Dhami, *Int. J. Comput. Appl*., **163(8)**, 15-19 (2017)
40. V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, G. Csányi, *Chem. Rev.*, **121(16)**, 10073-10141 (2021)
41. Q. Shi, M. Abdel-Aty, J. Lee, *Accid Anal Prev.*, **88**, 124-137 (2016)
42. M. A. Ahmadi, S. Zendehboudi, M. B. Dusseault, I. Chatzis, *Petroleum*, **2(1)**, 67-78 (2016)
43. D. Chen, Z. Pan, J. Liu, L. D. Connell, *Int. J. Coal Geol.*, **109**, 45-57 (2013)
44. N. M. Burnside, M. Naylor. *Int. J. Greenh. Gas Control*, **23**, 1-11 (2014)
45. Y. Chen, Q. Xie, A. Sari, P. V. Brady, A. Saeedi, *Fuel*, **215**, 171-177 (2018)
46. K. Lammers, R. Murphy, A. Riendeau, A. Smirnov, M. A. SchoonenD. R. Strongin, *Environ. Sci. Technol.,* **45(24)**, 10422-10428 (2011)
47. M. G. Rowan, *Regional geology and tectonics,* 205-234. (2020)