Measurement: Sensors xxx (xxxx) xxx

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

Density determination of CO2-Rich fluids within CCUS processes

ARTICLE INFO

Keywords:
CO2 density
Carbon capture
Utilization and storage
Machine learning
GPR
Transportation pipeline

ABSTRACT

Since accurate density determination is a crucial parameter for carbon dioxide (CO_2) fluid measurement and transportation pipeline designing, this study focuses on predicting the density of CO2-rich fluids in transportation pipelines for carbon capture, utilisation, and storage (CCUS) processes. In this regard, two supervised machine learning (ML) models, including the extra trees regressor (ETR) and Gaussian process regression (GPR), were developed using a diverse database of field-scale simulation samples. The models were compared based on their accuracy using statistical and graphical analysis. The GPR model performed better than the ETR model, achieving a smaller root mean square error (RMSE). The GPR model provides valuable insights for pipeline design, flow meter reliability, and uncertainty assessment in carbon storage. Experimental validation confirmed the robustness and practical applicability of the GPR model. This study demonstrates the potential of ML techniques to enhance the efficiency and reliability of CO2-rich fluid transportation in CCUS processes.

1. Introduction

In recent decades, global warming has become a major concern globally due to the significant increase in greenhouse gases, particularly CO2. Carbon capture utilisation and storage (CCUS) plays a crucial role in mitigating CO2 emissions by capturing and compressing CO2 at plants and transporting it through pipelines to subsurface reservoirs for storage. However, designing CO2 pipelines makes challenges due to operational variability, fluid phase variations, and impurities in CCUS processes, complicating fluid characterisation.

Efficient CO2 transport to geological storage sites requires well-designed pipelines considering fluid thermophysical properties. The variability in pressures, temperatures, and CO2-rich stream compositions affects fluid characterization, including density, phase behaviour, and viscosity which are critical for accurate fluid characterization in CCUS processes. Due to challenges and cost of density measurements, limited experimental data on CO2 fluids with impurities are available in the literature which necessitates new data collection to evaluate thermodynamic models [1].

On the other hand, modelling of CO2 density in CCUS processes is also complex due to variations of fluid's composition, process conditions, and data availability. Particular software, databases, and expert advice improve accuracy, while machine learning (ML) techniques provide an alternative for predicting fluid behaviour, especially in CO2-rich streams in pipelines.

Ali Mazari et al. [2] used different ML models including GPR and support vector regression (SVR) to predict CO2 thermophysical properties (e.g., CO2 solubility), and they found that GPR has the minimal error percentage compared to other models that were used to predict CO2 solubility among them. Also, Hung Vo Thanh a.et al. [3] used extreme gradient boosting (XGBoost), random forest (RF), and SVR to estimate both the solubility trapping index and residual trapping index in deep saline aquifers. XGBoost provided the most accurate prediction for CO2 trapping with $R^2 = 0.9998$ and RMSE = 0.0041, performing better than RF and SVR models. The XGBoost model offers a reliable, efficient approach for CO2 storage predictions. Hung Vo Thanh et al. [4] applied GPR, support vector machine (SVM), and RF to predict CO2 trapping efficiency in saline formations, and the results showed that GPR has the highest prediction accuracy ($R^2 = 0.992$, RMSE = 0.00491). These findings suggest that GPR, SVM, and RF can be effective ML models for estimating CO2 trapping performance in deep saline formations. ML techniques also show promising results for predicting CO2 mixture thermophysical properties, but there is a lack of precise density data for CCUS process due to limited experimental data.

Thorough validation is crucial for accurate predictions, particularly for diverse CO2 compounds, and comparing multiple ML models is vital for optimal CO2 density predictions [5,6]. ETR and GPR are popular for complex data of regression in CCUS. ETR performs better in predicting CO2 fluid properties, while GPR adapts to intricate data patterns without fixed parameters. ML models can generalize and scale their predictions of CO2 compound densities to diverse scenarios and pipeline configurations, eliminating the need for additional experiments. Therefore, our study uses ETR and GPR models for precise density prediction of CO2 and SO2 binary mixtures in transportation pipelines, aiding informed decisions in CCUS process.

2. Methods and procedures

In this section provides data collections and description, followed by explaining the ML algorithms used in this study, and concludes with the

This article is part of a special issue entitled: Supp: IMEKO 2024 published in Measurement: Sensors.

https://doi.org/10.1016/j.measen.2024.101739

methodology for creating predictive models based on ML methods.

2.1. Data preparation

Different data of density measurements were gathered from testing laboratories containing CO2 and SO2 binary mixtures at varying temperatures and pressures. Table 1 provides the details of the input and output parameters utilized in this study which have been used for developing the ML models.

2.2. Machine learning algorithms

We developed a reliable predictive model to compute the density of CO2 and SO2 binary mixtures by applying two powerful and robust ML methods, namely ETR and GPR to build a reliable prediction model. The following sections describe the theory of these ML methods in detail.

2.2.1. Extra trees regressor (ETR)

The ETR is known for its high level of accuracy, efficiency, and adaptability when performing regression tasks. This ensemble learning method excels in capturing intricate relationships within data and is an enhanced variant derived from the RF model [5,7].

The ETR performs better in capturing nonlinear relationships between input variables and CO2 density, making it particularly suitable for modelling complex thermodynamic properties. It is capable of effectively handling multiple input variables simultaneously, such as temperature, pressure, and composition, thereby facilitating comprehensive predictions of CO2 density in CCUS. Furthermore, it is appropriate for identifying and diagnosing measurement errors or anomalies that may be present in CCUS experimental or simulation data.

2.2.2. Gaussian process regression (GPR)

The GPR is a type of supervised learning technique that is employed to tackle regression and probabilistic classification problems. In the field of thermophysical properties, GPR is utilised to make predictions about properties like density, viscosity, and phase behaviour [8]. It plays a crucial role in accurately modelling heat transfer, fluid dynamics, and phase behaviour, thereby driving advancements in fields such as thermal management, chemical engineering, and phase behaviour calculations. By employing GPR, researchers and engineers are able to leverage a flexible and probabilistic framework that captures intricate relationships within thermophysical properties, providing valuable insights for both research and engineering applications.

2.3. Workflow of developing ML models

The ML model framework for predicting the density of CO2 and SO2 binary mixtures in CCUS transportation pipelines involves several key steps for model development as described in Fig. 1.

The development of the ML models starts with data collection and exploratory data analysis (EDA) which includes reviewing, analysing statistical, and pre-processing of features. Then, random subsampling is performed through data splitting to create training and testing subsets. Data splitting is an essential stage in ML model development. It involves dividing the database into training and testing subsets. The testing subset evaluates model performance, while the training subset builds knowledge bases. This dataset splitting boosts confidence in external predictions and shows the model's reliability across different datasets [9]. The data within the training subset were considered for the training model. ML methods are applied to the target data through specific algorithms and statistical models. Finally, the performance of each developed model is analysed and evaluated to identify the potential minimum error model among the two ML algorithms. The chosen model is then deployed to predict the CO2 and SO2 binary mixtures' densities at conditions similar to CCUS transportation.

In this study, the whole database was split into four training/testing subsets: 60 %–40 %, 70 %–30 %, 80 %–20 %, and 90 %–10 %. The initial

Table 1
Explanation of the inputs and outputs for ML models.

Properties	Unit	Description	Type	Min	Max
Temperature	K	Tunning temperature of CO2 and SO2 mixture	float	263.15	304.21
Pressure	MPa	Tunning pressure of CO2+So2 mixture	float	0.1	20.0
Phase Behaviour	–	It has two values, 1 refers to gas and 2 refers to liquid	integer	1	2
Experimental Density	(kg/m3)	Measuring Density base on temperature and pressure	float	1.80	1086.34

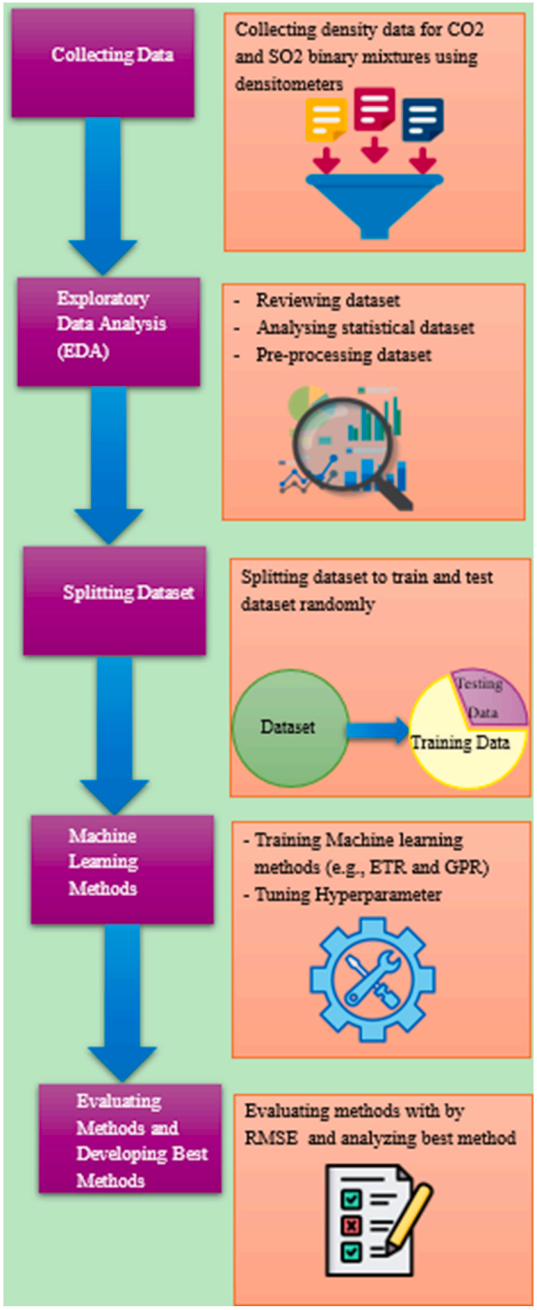

Fig. 1. The study's approach to create machine learning (ML) models.

subset was used for training, while the second one was designated for testing. However, the main challenge was related to selecting the most precise ML model. To overcome this, a statistical metric called RMSE was employed. RMSE is one of the criteria for evaluating regression problems [10]. It is shown using the following equation in (1):

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}} \quad (1)$$

Where y_i indicates CO2 compound density measured in the laboratory; $\hat{y}_i$ shows predicted density by ML method. N denotes the total number of instances used for evaluation.

3. Results and discussion

This section presents the prediction outcomes achieved using two machine learning techniques: ETR and GPR as previously described. It also discusses the evaluation of various hyperparameters (described below) and their influence on the model's performance.

3.1. Hyperparameter tuning

Hyperparameter is a parameter that controls the learning process in ML models. Hyperparameters are used to fine-tune model performance and control model complexity. The performance of a model is significantly impacted by its hyperparameters. This section explains the selection of optimal hyperparameters using a grid search method to prevent overfitting. In this study, we employed grid search to select the best tuning hyperparameter by assessing the RMSE value for each ML method. The specific tuning parameters for constructing the two predictive ML models can be found in Table 2. Training and testing were performed using the Scikit-learn Python package, which is a widely used open-source tool.

3.2. Performance evaluation of each ML methods

Four database subsets scenarios were analysed to determine the best ML method with the smallest RMSE. Table 3 displays the prediction outcomes of two studied ML models using the collected data samples to forecast the density of CO2 and SO2 binary mixtures, effectively. As demonstrated in *3*, for 60 %–40 % subset the developed methods showed achieving RMSE values of 1.749 and 3.550 by ETR and GPR, respectively which illustrate better performance of ETR method than GPR due to smaller value of RMSE ($RMSE_{ETR} = 1.749 < RMSE_{GPR} = 3.550$). However, the obtained RMSE values for the other subsets including 70 %–30 %, 80 %–20 %, and 90 %–10 % show different trends where the RMSE values of the GPR model were smaller than RMSE values of ETR Table 3.

By evaluating the overall performance of ML models across other data splitting, it becomes evident that in the three out of four data splitting subsets, the GPR technique demonstrates smaller prediction error compared to ETR according to statistical metrics. The best result of using GPR model for predicting density values obtained for the 90 %–10 % subset where RMSE calculated as 0.036 and 0.071 for training and testing subsets respectively. So, generally GPR performed better than ETR for both the training and testing subsets Table 3.

Fig. 2 visually represents the achieved accuracy using the 90 %–10 % subset by both models which has been used for predicting the density of CO2 and SO2 binary mixtures. The purpose of this function is to calculate performance metrics for both regression models on test data. First each model uses the input test data to make predictions output values for evaluating the model's accuracy and effectiveness on the test dataset. Second the performance metrics such as RMSE is calculated by comparing the predicted values with the test data values and The obtained number is calculated as a percentage and is considered as the accuracy of the model.

Table 2
The tuning parameter used to optimize the ML models.

model	parameter	Specific search range	optimal value
ETR	criterion	squared_error, absolute_error, friedman_mse	squared_error
	max_depth	1–3	10
	min_samples_split	1–3	2
	min_samples_leaf	1–3	1
	n_restarts_optimizer	190, 200, 220	5
GPR	kernel	1.0 * RBF(length_scale = length_scale) for length_scale in [0.1, 1.0, 10.0]	1**2 * RBF (length_scale = 0.1)
	alpha	0.1, 0.01, 0.001	0.001
	normalize_y	True, False	False
	n_restarts_optimizer	1–10	4

Table 3
The effectiveness of two ML models in predicting the density of CO2 and SO2 binary mixtures based on RMSE value for each database splitting, ML model, training subset and testing subset.

Dataset Splitting	Model	RMSE	
		training	testing
60 %–40 %	ETR	0.132	1.749
	GPR	5.344	5.358
70 %–30 %	ETR	0.193	2.046
	GPR	0.044	0.244
80 %–20 %	ETR	0.180	2.490
	GPR	0.035	0.197
90 %–10 %	ETR	0.193	1.459
	GPR	0.036	0.071

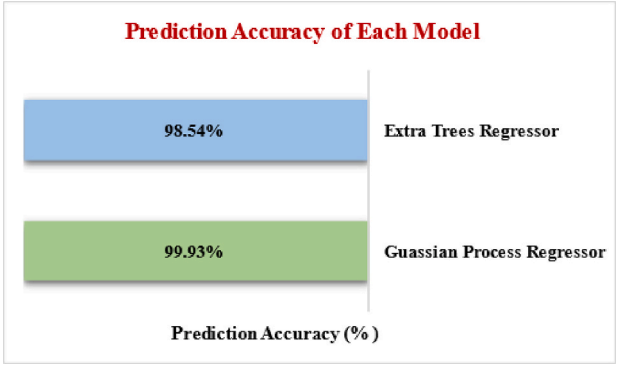

Fig. 2. To compare the performance criteria of each RMSE model, the prediction results of the models are calculated on the test data with two decimal places, converted to percentages, and then displayed in ascending sort on a horizontal plot.

This study found GPR over ETR, especially in the subset of 90 %–10 % which results in better accuracy of the predicted density values. The comparison of results also highlights the importance of selecting an optimal data splitting subset for ML performance in regression-based problems.

4. Conclusions

In this study, ML algorithms were used to estimate the density of binary CO2 and SO2 mixtures based on temperature, pressure and phase behaviour characteristics. This paper presents the potential application of two ML models named GPR and ETR algorithms in predicting the density of CO2-rich fluids in CCUS processes. Experimental data on CO2 and SO2 mixtures were collected to train and test ML models, where the

RMSE criterion was used to evaluate the performance of the developed ML models, and result of which showed that the GPR method performs better than the ETR in the density prediction which results in higher accuracy by fine-tuning hyperparameters. The developed methods showed appropriate prediction results, achieving appropriate RMSE values. This paper emphasized that regression-based methods often provide optimal performance. So, development of ML models can significantly improve estimating the values of various thermophysical properties within CCUS process e.g., density across a wider range of operational conditions with a focus on improving CCUS process performance.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] E. Luna-Ortiz, K. Szklarczyk-Marshall, M. Winter, E. McAllister-Fognini, Impact of hydrogen as impurity in the physical and transport properties of CO2 streams in CCS/CCUS transport systems: a technical discussion, in: Proceedings of the 15th Greenhouse Gas Control Technologies Conference, 2021, pp. 15–18.

[2] S.A. Mazari, et al., Prediction of thermo-physical properties of 1-Butyl-3-methylimidazolium hexafluorophosphate for CO2 capture using machine learning models, J. Mol. Liq. 327 (2021) 114785.

[3] H.V. Thanh, Q. Yasin, W.J. Al-Mudhafar, K.-K. Lee, Knowledge-based machine learning techniques for accurate prediction of CO2 storage performance in underground saline aquifers, Appl. Energy 314 (2022) 118985.

[4] H.V. Thanh, K.-K. Lee, Application of machine learning to predict CO2 trapping performance in deep saline aquifers, Energy 239 (2022) 122457.

[5] U. Ashraf, et al., Classification of reservoir facies using well log and 3D seismic attributes for prospect evaluation and field development: a case study of Sawan gas field, Pakistan, J. Pet. Sci. Eng. 175 (2019) 338–351.

[6] V. Venkatraman, B.K. Alsberg, Predicting CO2 capture of ionic liquids using machine learning, J. CO2 Util. 21 (2017) 162–168.

[7] J. Kondori, M.I. Miah, S. Zendehboudi, F. Khan, D. Heagle, Hybrid connectionist models to assess recovery performance of low salinity water injection, J. Pet. Sci. Eng. 197 (2021) 107833.

[8] A.A. Sulaimon, A.R.T. Salang, A. Qasim, S.A. Akintola, C.D.A.P. Wifred, Screening of ionic liquids for co2 capture using data analytics techniques, Platf. A J. Eng. 7 (2) (2023) 1–12.

[9] W. J Al-Mudhafar, Integrating lithofacies and well logging data into smooth generalized additive model for improved permeability estimation: zubair formation, South Rumaila oil field, Mar. Geophys. Res. 40 (2019) 315–332.

[10] D.S.K. Karunasingha, Root mean square error or mean absolute error? Use their ratio as well, Inf. Sci. 585 (2022) 609–629.

Shirin Gholami[a,*], Elahe Rostaminikoo[b], Leila Khajenoori[b], Hamid Reza Nasriani[b]

[a] *School of Computer Sciences, Azad University, Iran*

[b] *School of Engineering and Computing, University of Central Lancashire (UCLan), UK*

[*] Corresponding author.

E-mail address: gholamishirin06@gmail.com (S. Gholami).