

Towards More Accurate and Explainable Supervised Learning-Based Prediction of Deliverability for Underground Natural Gas Storage

Aliyuda Ali^{a,*}, Kachalla Aliyuda^b, Nouh Elmitwally^a, and Abdulwahab Muhammad Bello^c

^a School of Computing and Digital Technology, Birmingham City University, B4 7XG, Birmingham, United Kingdom

^b Department of Geology and Petroleum Geology, University of Aberdeen, AB24 3UE, Aberdeen, United Kingdom

^c Centre for Integrative Petroleum Research, College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

Abstract

Numerous subsurface factors, including geology and fluid properties, can affect the connectivity of the storage spaces in depleted reservoirs; hence, fluid flow simulations become more complicated, and predicting their deliverability remains challenging. This paper applies Machine Learning (ML) techniques to predict the deliverability of underground natural gas storage (UNGS) in depleted reservoirs. First, three baseline models were developed based on Support Vector Regression (SVR), Artificial Neural Network (ANN), and Random Forest (RF) algorithms. To improve the accuracy of the RF model as the best-performing baseline model, a unified framework, referred to as SARF, was developed. SARF combines the capabilities of Sparse Autoencoder (SA) and that of Random Forest (RF). To achieve this, the internal representations of the SA, which constitute extracted features of the input variables, are used in RF to develop the proposed SARF framework. The predictive capabilities of the baseline models and the proposed SARF model were validated using 3744 real-world storage data samples of 52 active storage reservoirs in the United States. The experimental result of this study shows that the proposed SARF model achieved an average 5.7% increase in accuracy on four separate data partitions over the baseline RF model. Furthermore, a set of eXplainable Artificial Intelligence (XAI) methods were developed to provide an intuitive explanation of which factors influence the deliverability of reservoir storage. The visualizations developed using the XAI method provide an easy-to-understand interpretation of how the SARF model predicted the deliverability values for separate reservoirs.

Keywords— Artificial neural network, data-driven modeling, interpretable machine learning, natural gas industry, random forests, support vector regression.

List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
AQI	Air Quality Index
BG	Base Gas
CNN	Convolutional Neural Network
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
EIA	Energy Information Administration
FERC	Federal Energy Regulatory Commission
IEA	International Energy Agency
MAE	Mean Absolute Error
Mcf	Million cubic feet
ML	Machine Learning
MSE	Mean Squared Error
PDD	Probability Density Distributions
RF	Random Forest
RMSE	Root Mean Squared Error
SA	Sparse Autoencoder
SARF	Sparse Autoencoder Random Forest
SHAP	SHapley Additive exPlanations
SOMs	Self-Organizing Maps
SVM	Support Vector Machines
SVR	Support Vector Regression
TFC	Total Field Capacity
UNGS	Underground Natural Gas Storage
USA	United States of America
WGC	Working Gas Capacity

1. Introduction

According to the 2022 gas market reports by the International Energy Agency (IEA), worldwide natural gas consumption rebounded by 4.6% in 2021 [1]. This rebounding in consumption is more than double the decline observed in 2020. Economic recovery and extreme weather were among the factors that led to strong demand growth for natural gas in 2021. As a result, the natural gas supply did not keep pace, which, coupled with unanticipated outages, resulted in tight markets and abrupt price increases. UNGS reservoirs play a vital role in ensuring that any excess gas delivered during the low-demand season is accessible to meet seasonal consumption demands and serve as a security against unexpected supply interruptions [2], [3]. Natural gas can be stored in various ways. It is most commonly stored in the subsurface, under pressure in three types of formations: depleted reservoirs, aquifers, and salt caverns [4], [5]. Each of these storages has its own physical and economic characteristics, which govern its feasibility for particular applications. These characteristics include maintenance cost, retention capability, cycling capability, deliverable rates, and site preparation [6]. Depleted reservoirs (oil and/or gas reservoirs) are subsurface formations whose recoverable natural resources have already been exhausted and, afterward, converted to natural gas storage facilities. Geologically, for a depleted reservoir to be converted into an underground natural gas storage reservoir, it must have some of the critical elements of a good hydrocarbon play such as good porosity and permeability, presence of good seal rock, presence of cap rock, presence of trap and absence of open fault/leakages. Hence, their already known geological properties make them suitable for storing natural gas effectively. These properties make depleted reservoirs cost-effective in development, operation, and maintenance compared to salt caverns and aquifers [7]. Reconditioning depleted reservoirs from production to storage facilities benefits from using already developed reservoirs with existing equipment, and pipeline connections left when the reservoirs are productive [7]. Depleted reservoirs have demonstrated their geological suitability, keeping their trapped hydrocarbon accumulations for millions of years. Natural gas storage reservoirs in the USA are monitored regularly by Federal Energy Regulatory Commission (FERC). Regular assessments are carried out by the FERC and the operating companies of the reservoirs to consider reservoir properties, overfilling, well design, over pressuring and overburden geology in order to identify potential migration pathways/leaks. Numerous interconnected pore spaces found within the storage sites of depleted reservoirs can complicate fluid flow predictions [8]; thus, forecasting their deliverability is challenging. The use of depleted reservoirs to store natural gas has been considered a strategic practice to balance supply and demand in various parts of the world [9]. The first successful usage of depleted underground reservoirs to store natural gas was reported in 1915, in Ontario, Canada [10]. Since then, a multitude of such reservoirs have been developed in the Middle East, North America, Asia-Oceanic, Europe, and other parts of the world [7].

In the energy industry, energy conservation is one of the crucial issues of the 21st century [11]. Recently, the application of AI and ML approaches to provide solutions to complex engineering problems has received considerable attention in the energy industry [12], [13], [14], [15], due to their successful application in various areas, including, electricity demand and consumption forecasting [16], [17], [18]; controlling room temperature to minimize electricity cost [19]; forecasting building energy consumption [20]; providing explicit demand response from domestic boilers [21]; and, evaluating energy efficiency parameters [22]. Despite the successful application of AI and ML approaches in the energy industry, there are comparatively fewer studies utilizing AI and ML techniques to predict the deliverability of natural gas storage in depleted reservoirs.

It is very helpful to distinguish between natural gas reservoirs and natural gas storage reservoirs. Natural gas reservoirs are underground geological formations that currently produce natural gas from their recoverable natural resources, whereas, underground natural gas storage reservoirs are depleted natural gas reservoirs whose recoverable hydrocarbon deposits have been exhausted and, afterward, converted to natural gas storage facilities. There has been a wealth of data collected from numerous UNGS sites developed in various parts of the world, from 1915 to date, providing great scope for the application of developmental research using AI and ML techniques. Therefore, the main aim of this study is to examine the capability of ML algorithms in predicting the deliverability of UNGS in depleted reservoirs using real-world data of active storage reservoirs.

In this study, we propose a unified SARF framework that combines the capabilities of SA and RF techniques to improve the prediction of deliverability of UNGS in depleted reservoirs. First, three baseline models based on SVR, ANN, and RF algorithms were developed for comparison. Then, a SA was trained using the input variables to learn and extract significant features. The extracted features were then used by the best-performing baseline model, in this case, the RF model, to make the final predictions. Thus, the proposed unified SARF model is hypothesized to exhibit superior prediction performance over the standalone RF model and the other baseline models. 3744 real-world data samples from 52 active storage reservoirs operating in 7 States of the USA were

used to validate the effectiveness of the proposed SARF model. The aim of this study is to predict the deliverability of UNGS in depleted reservoirs using ML techniques, with the specific objectives to:

1. Examine the capabilities of ML algorithms in predicting the deliverability of UNGS in depleted reservoirs.
2. Propose a new framework (SARF) that unifies the capabilities of SA and RF algorithms to improve the predictive capability of RF algorithm.
3. Validate the ML algorithms and the proposed SARF framework using real-world data from active depleted reservoirs.
4. Outline the first-time use of XAI for interpreting and explaining ML model for deliverability prediction by providing visual interpretations of why a model arrived at predicting a particular value.

Unsupervised learning techniques have significantly improved the performance of supervised ML algorithms. Some previously reported studies that utilized the strength of SA and other unsupervised learning techniques to improve the performance of supervised ML algorithms are presented in Table 1.

Table 1. Summary of works that used unsupervised learning techniques to improve the performance of supervised ML algorithms.

Reference	Year	Underlying AI/ML Technique	Outcomes & Key Results
M. Al-Qatf, <i>et al.</i> [23]	2018	SA combined with SVM	A unified framework for network intrusion detection. Results showed that integrating SA into SVM has improved both binary and multiclass classifications accuracy over previously reported approaches.
I. D. Mienye, <i>et al.</i> [24]	2020	SA combined with ANN	A two-stage method that improves the prediction of heart disease was developed. Results revealed that the two-stage method performs better than standalone ANN and other related scholarly works.
H. Sewani and R. Kashef. [25]	2020	SA combined with CNN	A unified method for efficiently diagnosing autism spectrum disorder in children was developed. Results showed that the proposed method exhibits superior performance compared to other existing methods.
H. Liu and R. Yang [26]	2021	SA coupled with Multi-Objective Wolf Colony algorithm.	An effective model for the prediction of Air Quality Index (AQI) that can provide a reliable assurance for air quality warnings and pollution control was developed. Results revealed that the proposed model perform better than other comparative models.
B. Scheibel, <i>et al.</i> [27]	2021	DBSCAN clustering technique combined with a novel distance metric and an iterative parameter setting algorithm.	An integrated framework that demonstrates the use and design of engineering drawings in an automated real-world manufacturing process.
P. Calvo-Bascones, [28]	2021	K-means and Self-Organizing Maps (SOMs) algorithms combined with PDD.	A unified framework that facilitates anomalies detection and surveillance of industrial component behaviors. The proposed approach was applied to hydropower plant to monitor the bearing temperature of a turbine.
H. Wu. [29]	2021	SA combined with Logistic Regression-Recursive algorithm.	A deep learning-based hybrid approach for cancer diagnosis was developed. Results showed that the proposed hybrid approach performed better than other comparison techniques.

Previous work by the authors [30], reported the first study that applied ML algorithms to predict the deliverability of UNGS in salt caverns. In that work, the capabilities of three ML techniques namely, ANN, SVM, and RF were investigated. Experimental results showed that the RF algorithm performs better than the ANN and SVM on

different data partitions. It is important to note that the storage areas in salt caverns are characterized by a single large open space which is used to store or withdraw natural gas without contracting the resistance of the pore spaces to the natural gas flow. On the contrary, storage space in depleted reservoirs is characterized by numerous interconnections of pore spaces whose resistance to natural gas flow causes contraction. This makes deliverability prediction of UNGS in depleted reservoirs more challenging than in salt caverns. Furthermore, in terms of storage capacity, depleted reservoirs have the highest percentage of storage worldwide in comparison to salt caverns. Thus, as an extension to previous work [30], this study proposes a hybrid approach that combines the capabilities of SA and RF to predict the deliverability of UNGS in depleted reservoirs using real-world data.

The remaining part of this paper is organized as follows. Section two presents brief theoretical descriptions of the three ML algorithms used as the baseline models. It is followed by the development process of the baseline ML models. Section three presents the workflow and the processes of developing the proposed SARF framework. Experimental results and discussion are presented in section four. Section five demonstrates the use of XAI technique known as SHAP to interpret the SARF model's working principle. Conclusion and direction for future work are presented in section six.

2. The Baseline Machine Learning Models

This section starts with brief theoretical descriptions of the three ML algorithms employed to develop the baseline models: SVR, ANN, and RF techniques. This is followed by the model development process for the three baseline algorithms.

2.1. Support Vector Regression (SVR)

The SVR algorithm is a variant of SVMs that is used to solve nonlinear regression problems [29]. Given a training set $T = \{(x_1, y_1), \dots, (x_l, y_l)\}$, where $x_i \in R^n, y_i \in y = R, i = 1, \dots, l$; the starting point of SVR is to find a decision function $f(x) = (w^T \cdot x) + b$, by introducing the transformation $x = \Phi(x)$ and the corresponding kernel $K(x_1, x_2) = (\Phi(x_1) \cdot \Phi(x_2))$. SVR constructs and solves a convex quadratic programming problem to obtain the optimal regression function. The utilization of the nonlinear kernel function is the key in SVR applications. In SVR, the nonlinear kernel function maps nonlinear data set into a high-dimensional linear space for the optimization process to obtain a nonlinear regression model. In this work, the Gaussian kernel function type is used and can be expressed mathematically as:

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2), \text{ for } \gamma > 0. \quad (1)$$

The Gaussian kernel was chosen as the kernel function for the SVR method due to its computational efficacy, simplicity of adaptation for optimizing other adaptive methods, and its ease of use for dealing with complex and nonlinear parameters [31].

2.2. Artificial Neural Network (ANN)

Similar to the synapses in a human brain, information in ANN is processed by connecting different simple nodes to establish complex networks [32]. In ANN, each node receives signals as input, processes it using an activation function and passes its output to other nodes through a weighted connection. Thus, the output of an ANN is determined by the architecture of the network (network structure), the weight value, and the kind of activation function used. A schematic of a single artificial neuron is depicted in Figure 1. Let consider a unit of an artificial neuron i , as depicted in Figure 1, suppose its input signals are connected from other units say x_i (where $i = 1, 2, \dots, n$) with corresponding weights w_i , to process the unit, two key operations are performed namely, summation and activation of input signals.

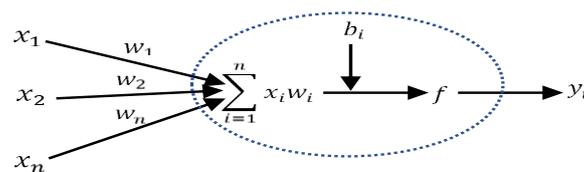


Fig. 1. A schematic of an artificial neuron process.

The output unit y_i is defined as:

$$y_i = f(\sum_{i=1}^n x_i w_i + b_i) \quad (2)$$

where b_i denotes a bias for the input i , and f denotes the activation function. In this work, the sigmoid function is used and can be expressed as

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

2.3. Random Forest (RF)

The RF is a method that employs the capability of ensemble learning and decision trees to generate accurate predictive models. The technique is suitable for handling a degree of nonlinearity in the data by providing deep insights into data structure [33]. Here, ensemble learning refers to a type of ML whose main aim is to develop a robust predictive model by combining the predictions of multiple base models. The robust predictive model is designed to perform better than any contributing ensemble member.

The following steps summarize the RF algorithm for achieving regression task:

Step 1: From the original data, produce bootstrap samples of n_{tree} .

Step 2: Build an untrimmed regression tree for each of the produced bootstrap samples.

Step 3: At each knot, randomly sample n_{try} of the predictors and pick from among the variables the best split.

Step 4: Generate prediction of new data by combining the predictions of the n_{tree} .

2.4. The Baseline Machine Learning Model Development Process

Developing the baseline ML models for predicting the deliverability of UNGS in depleted reservoirs requires many steps as depicted in Figure 2. The process begins with pre-processing the entire input/output data. This was achieved by normalizing the entire input/output data to have values between 0 and 1. Next, the normalized dataset was randomly partitioned into training and testing sets. In this work, the input/output data were partitioned into four different sets with different proportions: 90% - 10%, 80% - 20%, 70% - 30%, and 60% - 40% where the first and second proportions stand for the training and testing data sets, respectively. The reason for partitioning the input/output data into four various partitions is to examine the performance and consistency of the baseline models on different data partitions. Next, the three baseline ML algorithms are implemented to develop regression models for the prediction task. This involves feeding each learning algorithm with the training data sets for each data partition. To evaluate the performance of the baseline models, three performance metrics were selected, namely, coefficient of determination (R^2), root mean squared error (RMSE) and mean absolute error (MAE). The R^2 is selected to evaluate the prediction accuracy and is expressed as follows [34]:

$$R^2 = \left[1 - \frac{\frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2}{var(y)} \right] \quad (4)$$

where N denotes the total number of data samples, \hat{y}_t and y_t denote the actual and predicted values at time t , respectively. The R^2 is a statistical quantity that indicates how well a model's outcome matches the actual outcome. Its values range from 0 to 1. The lower the R^2 value, the poorer the prediction accuracy; the higher the value, the better the prediction accuracy. The RMSE is the square root of the average sum of deviation of predicted values and actual values. The RMSE is selected to evaluate the prediction error and is expressed as follows [35]:

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{t=1}^N (\hat{y}_t - y_t)^2} \quad (5)$$

The RMSE has values from 0 and above. An RMSE value of 0 indicates a perfect prediction without error, whereas the farther the RMSE value from 0, the larger the prediction error. The MAE is a measure of errors between actual and predicted values, which is calculated as the sum of absolute errors divided by the sample size. The MAE is selected to evaluate the model prediction error further and is expressed as follows:

$$MAE = \frac{1}{N} \sum_{t=1}^N |\hat{y}_t - y_t| \quad (6)$$

where $|\hat{y}_t - y_t|$ represents the absolute errors between the actual and predicted values for all the data samples.

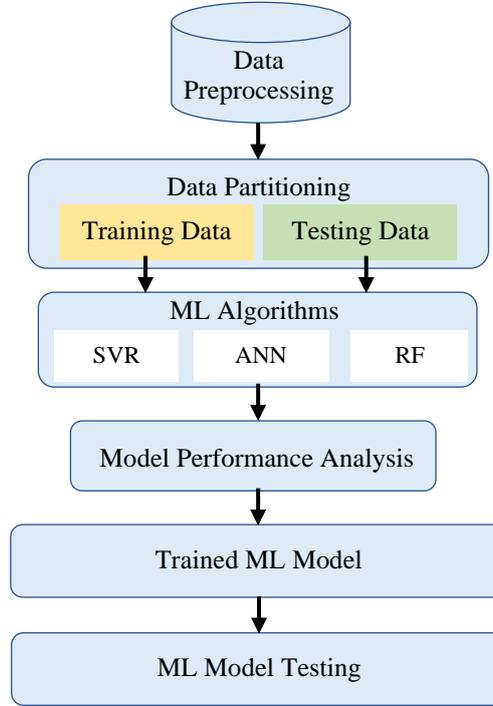


Fig. 2. Baseline ML model development process to predict the deliverability of UNGS in depleted reservoirs.

After training and analyzing the performance of the baseline models, the best-trained model for each algorithm is retained. The best-trained model for each algorithm is tested using the testing sets to predict the deliverability of UNGS in depleted reservoirs. In training ML models, hyperparameter optimization improves model performance by determining the optimal values for a given model. In this work, the Grid Search technique was used to obtain the optimal values for both the baseline ML and the proposed SARF models. The technique was implemented using GridSearchCV (also known as GridSearch cross-validation) function, provided by the Scikit-learn library. Even though the Grid Search technique might be time-consuming when dealing with a large dataset with many input variables, the major reason for choosing it is to benefit from its capability to generate the best candidates from a grid of parameter values exhaustively. Table 2 gives a detailed description of parameter values used for developing the baseline models.

Table 2. Parameters used for each baseline ML model.

Machine Learning Algorithm	Input variables	Output variable	Hyperparameters
ANN	Base gas Working gas capacity Total field capacity	Deliverability	Hidden layers: 1, neurons: 10, training function: Levenberg-Marquardt, activation function: Tangent sigmoid, performance function: mean squared error, initial training gain: 0.001, training gain decrease factor: 0.1, training gain increase factor: 10, maximum training gain: 1e10, minimum performance gradient: 1e-7, maximum number epochs: 500.

SVR			Kernel function: Gaussian, kernel scale: 3.7, box constrain: 0.741, Epsilon: 0.074, solver: Sequential Minimal Optimization (SMO).
RF			Regression trees: 200, predictor selection: interaction-curvature, method: Bootstrap aggregating, learning cycles: 200.

3. The Proposed SARF Framework

This section begins with a brief theoretical description of SA, followed by the proposed SARF model development process. The section ends with the data source, description, and pre-processing.

3.1. Sparse Autoencoder (SA)

An autoencoder is a ML technique that consists of an encoder and a decoder. It is an unsupervised ML and dimensionality reduction algorithm that automatically learns from unlabelled data and sets the output values to be the same as the input values. The objective of training an autoencoder is to learn and extract the representations of the input data, which are then used to reconstruct the original input data. Suppose we have an input to the autoencoder as a vector x , the encoder maps the input vector to a new representation vector l , which is then decoded to reconstruct the original input as \hat{x} . Mapping the input data to a new representation is done by the encoder and reconstructing the original input data using the new representations is done by the decoder as expressed in (7) and (8).

$$L = g(Wx + b), \quad (7)$$

$$\hat{X} = h(\widehat{W}l + \widehat{b}). \quad (8)$$

In the above equations, g and h represent the activation functions for the hidden and output layers. W and \widehat{W} represent the weight matrices, and b and \widehat{b} represent the bias vectors for the encoder and the decoder, respectively. In this work, the log-sigmoid activation function, as shown in equation (9) was used to establish the new representations.

$$g = h = \frac{1}{1+e^{-x}} \quad (9)$$

Unlike traditional autoencoders, this works used a sparse autoencoder (SA) to obtain the new representations of the input variables. Thus, a regularizer was added to the cost function to stimulate sparsity of the autoencoder. This regularizer computes the activated mean output of the neurons in the hidden layer. Let $\widehat{\rho}_i$ be the activated mean output of the neurons in the hidden layer as expressed in (10),

$$\widehat{\rho}_i = \frac{1}{N} \sum_{j=1}^N L_i(x_j) \quad (10)$$

here, N stands for the number of training samples, i , stands for i th neuron, and x_j stands for the j th training samples. The mean activation output $\widehat{\rho}_i$ approaches ρ that is a value close to zero. Thus, to achieve sparsity, Kullback-Leibler (KL) divergence is introduced as shown in equation (11) and used later to add the regularizer to the cost function.

$$\Omega_{sparsity} = \sum_{i=1}^k \rho \log \left(\frac{\rho}{\widehat{\rho}_i} \right) + (1 - \rho) \log \left(\frac{1 - \rho}{1 - \widehat{\rho}_i} \right) \quad (11)$$

Here, k denotes the total number of neurons in a layer, and ρ denotes sparsity proportion, which is the required activation value. A neuron with a high output activation value is considered to be 'firing'. When a term that constrains the values of the activated mean output of the neuron is added to a cost function, the autoencoder is then stimulated in a manner that allows the neurons in the hidden layer to fire to a low number of training samples. In training a SA, an increase in the values of the weights and a decrease in the values of biases can lead to a possibility of making the sparsity regularizer small. Therefore, to prevent such from happening, a regularization

term L_2 was added to the cost function. Introducing this L_2 -norm as a penalty to the cost function improves the generalization ability of the autoencoder and makes the solution “smoother”. In this work, mean squared error (MSE) was used as a cost function to train the SA. Thus, the SA cost function E comprises the MSE, L_2 regularization term, and sparsity regularization term, as shown in (12)

$$E = \frac{1}{N} \sum_{n=1}^N \sum_{d=1}^D (x_{dn} - \hat{x}_{dn})^2 + \gamma * L_2 + \beta * \Omega_{sparsity} \quad (12)$$

where γ and β represent the coefficients of L_2 and $\Omega_{sparsity}$, respectively.

3.2. The Proposed SARF Model Development Process

A flowchart that describes the stages and processes of developing the proposed SARF framework for deliverability prediction of UNGS in depleted reservoirs is shown in Figure 3.

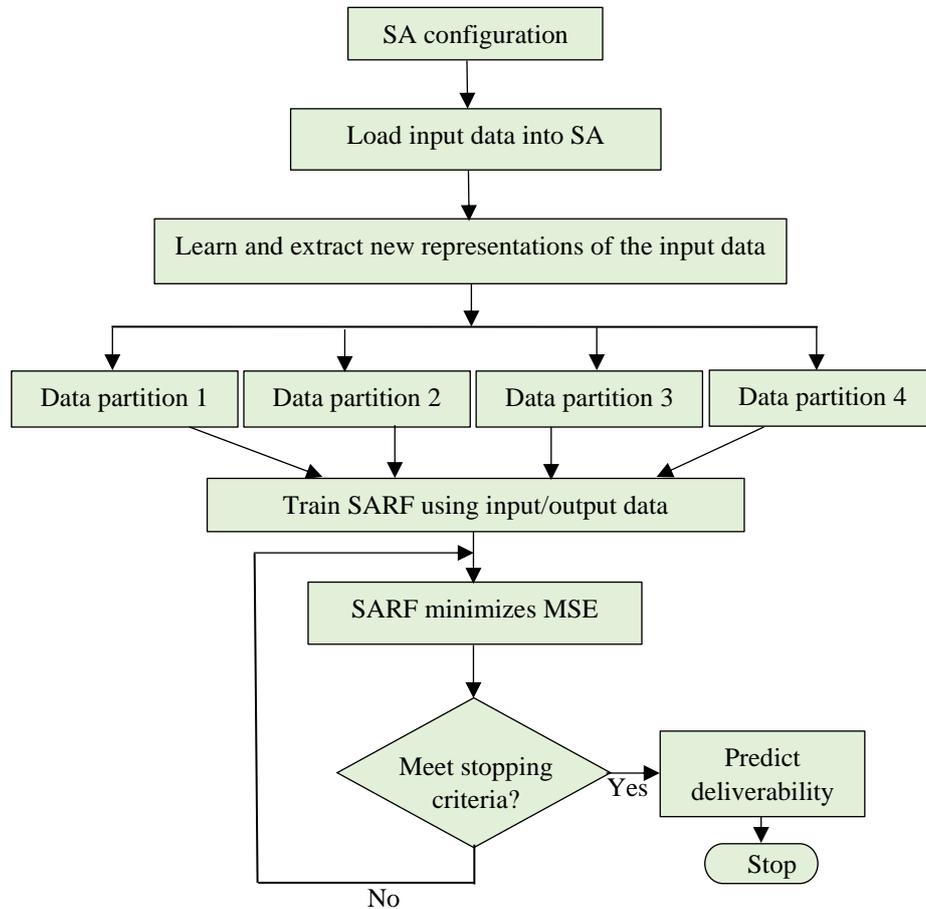


Fig. 3. The workflow of the proposed SARF model development for deliverability prediction of UNGS in depleted reservoirs.

As shown in Figure 3, the development process starts with configuring a SA using the required parameters. It learns and extracts significant features from the input data and stores them as new representations. The proposed SARF used the same data partition as the baseline model. Next, training data sets for each data partition were fed into an RF algorithm to train the proposed SARF model. Training the SARF model is to minimize the MSE between the actual deliverability value and the predicted deliverability value using the new representation features as input until the stopping criteria is met. In this work, the stopping criteria is met when the maximum number of epochs is reached. Finally, the trained model is used to predict the deliverability using the testing data. As in the case of developing the baseline models, R^2 , RMSE and MAE were used to evaluate the prediction accuracy and errors of the proposed SARF framework. Also, the Grid Search technique was used to get the optimal parameter values for SA algorithm that forms the SARF framework. Table 3 gives a detailed description of the SA parameter values used for developing the proposed SARF model. Figure 4 shows a flow chart that describes how SHAP method is integrated into the proposed SARF model for predicting deliverability.

Table 3. SA parameters used for developing the proposed SARF model.

Algorithm	Hyperparameters
Sparse Autoencoder	Number of encoders: 1, sparsity regularization: 4.000, sparsity proportion: 0.1500, L2 weight regularization: 0.001, encoder activation function: log-sigmoid, training algorithm: scaled conjugate gradient, cost function: mean squared error Sparse.

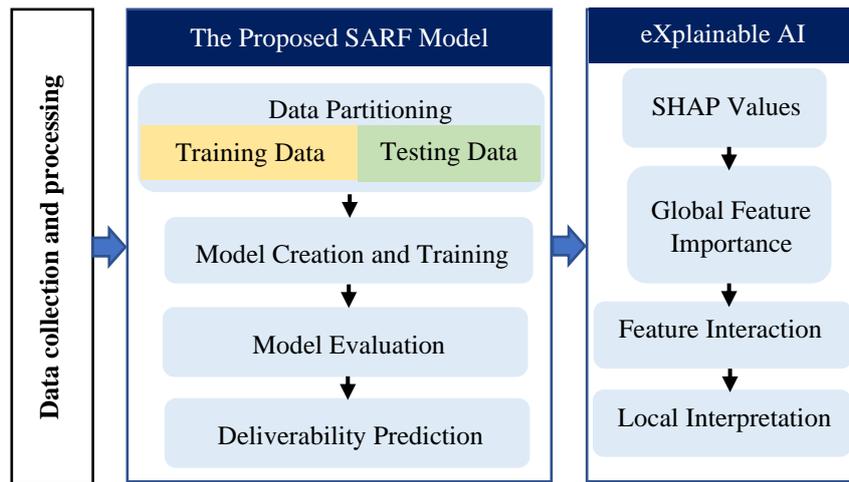


Figure 4. A flow chart describing the integration of SHAP method into the proposed SARF model for deliverability prediction.

3.3. Data Collection, Description, and Pre-processing

The dataset used was collected from the official site of the US Energy Information Administration (EIA) [36]. The dataset consists of 3744 samples of monthly (January to December) field storage for six years (2016 through to 2021) of 52 active depleted reservoirs operating in 7 States of the USA. The field storage data are collected from the operating companies by the EIA and are drawn to the EIA-191 Field Level Storage Data Form, which is published on the EIA official website. Figure 5 presents the number of depleted reservoirs in each of the 7 States whose data were used in this study. The State abbreviations AL, AR, CA, CO, KS, LA, and VA shown in figure 5 stand for Alabama, Arkansas, California, Colorado, Kansas, Louisiana, and Virginia, respectively.

To select the appropriate predictors for forecasting the deliverability of UNGS in depleted reservoirs, information on the factors that govern the deliverability of a well in a natural gas reservoir is essential. As described in [30], [37] and [38], the fundamental working behavior of underground depleted reservoir storage involves reconditioning the reservoir prior to storage (gas injection) to create an adequate underground pressurized space, as in the case of newly drilled oil or gas wells, the higher the pressure in the depleted reservoir storage, the more readily natural gas may be withdrawn. Once the pressure drops to below that of the wellhead, there will be no pressure differential needed to drive the natural gas out of the depleted reservoir storage.

Several studies have examined the relationship between geological/engineering factors with reservoir performance metrics such as reservoir daily production rate, depletion rate, and recovery factor. The authors in [39] and [40] reported a strong relationship between reservoir performance and reservoir volumes as well as size of the field or reservoir. Similarly, [41] suggested that working gas, base gas, and total gas are the key factors that control the deliverability of underground natural gas storage. Hence, the decision to use them as input variables. From the EIA monthly storage data for the 52 depleted reservoirs presented in Figure 5, deliverability, which is referred to as the measure of the amount of natural gas that can be delivered to the marketplace from a depleted reservoir and is measured in Mcf/day, is defined as the dependent variable (output) whereas base gas (BG), working gas capacity (WGC), and total field capacity (TFC), are defined as independent variables (inputs).

Description of the real-world gas storage deliverability data used in this study is presented in Table 4, and a statistical description of the input/output variables in terms of their minimum, maximum and average values is given in Table 5. The entire data set was normalized to range between 0 and 1 to preserve feature variations, enhance convergence time, and improve model stability, before feeding it into the ML algorithms.

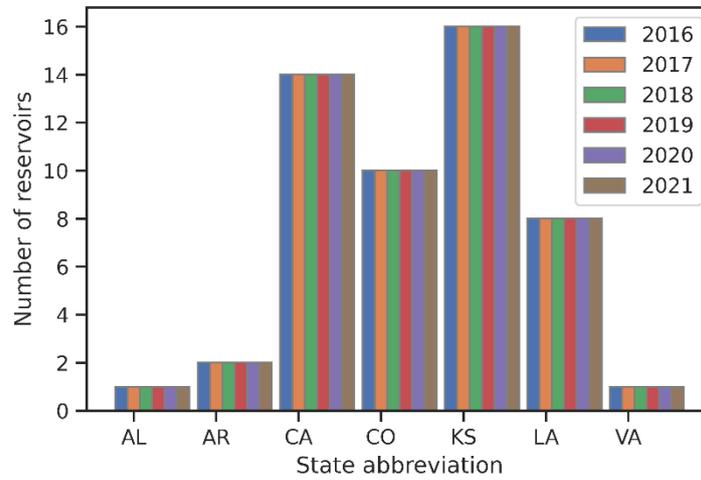


Fig. 5. UNGS in depleted reservoirs whose data were used in this paper.

Table 4. Description of all variables used in this study

Variable (Mcf)	Description
TFC	The highest volume of natural gas that can be stored in an underground storage infrastructure in agreement with its design, which includes the reservoir characterization, installed facility and working process in the specific place.
BG	The volume/proportion of natural gas that must remain in the storage reservoir as permanent inventory to maintain adequate pressure and rates of delivery throughout the withdrawal season.
WGC	Represents total field capacity minus base gas.
Deliverability	The measure of the quantity/amount of natural gas that can be withdrawn (delivered) from the depleted storage reservoir.

Table 5. The statistical description of the input/output variables used in this study.

Variable (Mcf)	Minimum	Maximum	Average
BG	38104	99492767	13483757.79
WGC	160000	86200000	16818890.91
TFC)	320340	167725000	30383846.46
Deliverability	500	1860000	327927.86

4. Experimental Results and Discussion

This section presents and discusses the results of both the baseline models and the proposed SARF models.

4.1 The Baseline Models

Experimental results for each baseline model on four data partitions for training, testing, and overall performance were computed and presented in Table 6.

Table 6. Comparison of baseline ML models' performance in predicting the deliverability of UNGS in depleted reservoirs using various data partitions.

Data partition	Model	Training			Testing			Overall		
		R ²	MAE	RMSE	R ²	MAE	RMSE	R ²	MAE	RMSE
90% - 10%	SVR	0.8686	0.0064	0.0803	0.8654	0.0063	0.0794	0.8680	0.0064	0.0802
	ANN	0.8862	0.0054	0.0738	0.8468	0.0085	0.0921	0.8816	0.0057	0.0758
	RF	0.9356	0.0032	0.0562	0.9243	0.0043	0.0658	0.9342	0.0033	0.0573
80% - 20%	SVR	0.8698	0.0063	0.0794	0.8639	0.0070	0.0837	0.8685	0.0064	0.0803
	ANN	0.8751	0.0060	0.0773	0.8709	0.0065	0.0806	0.8742	0.0061	0.0780
	RF	0.9438	0.0028	0.0527	0.9417	0.0030	0.0552	0.9434	0.0028	0.0532
70% - 30%	SVR	0.8720	0.0063	0.0793	0.8772	0.0060	0.0775	0.8735	0.0062	0.0788
	ANN	0.8800	0.0058	0.0764	0.8782	0.0059	0.0768	0.8786	0.0059	0.0767
	RF	0.9389	0.0031	0.0554	0.9333	0.0033	0.0577	0.9350	0.0032	0.0570
60% - 40%	SVR	0.8738	0.0061	0.0783	0.8604	0.0070	0.0837	0.8684	0.0065	0.0805
	ANN	0.8582	0.0069	0.0830	0.8552	0.0068	0.0823	0.8571	0.0068	0.0827
	RF	0.9278	0.0036	0.0603	0.9238	0.0037	0.0609	0.9261	0.0037	0.0605

The results shown in Table 6 reveal that each of the three baseline models developed in this work achieved an accuracy of 85% and above with respect to training, testing, and overall performance on all four data partitions. It can be observed, in terms of individual model performance on each data partition, from the text highlighted in bold. The RF model recorded the highest prediction accuracy on all data partitions with prediction accuracies of 93%, 94%, 93%, and 92% on 90% - 10% data partition, 80% - 20% data partition, 70% - 30% data partition, and 60% - 40% data partition, respectively. The highest prediction accuracy of the RF model is 94%, and this was achieved on data partition 2 (80% training - 20% testing). SVR model achieved its highest prediction accuracy of 87% on data partition 3 (70% training - 30% testing), whereas the ANN model achieved its highest prediction accuracy of 88% on data partition 1 (90% training - 10% testing). With respect to bias minimization and consistency between training and testing outputs, it can be observed that the RF model recorded the best outputs on all four data partitions with insignificant variations between the training and testing outputs over SVR and ANN models.

4.2 The Proposed SARF Model

RF algorithm performed better over ANN and SVR; however, the highest accuracy of 94% recorded by RF is unsatisfactory, which requires further improvement. As stated in section 1, the proposed SARF model aims to improve the predictive capability of the RF algorithm. The proposed SARF model was applied to the four data partitions to evaluate its performance. Results of the proposed SARF model are presented in Table 7.

Table 7. Results of SARF model performance on predicting the deliverability of UNGS in depleted reservoirs using various data partitions.

Data partition	Training			Testing			Overall		
	R ²	MAE	RMSE	R ²	MAE	RMSE	R ²	MAE	RMSE
90% - 10%	0.9941	0.0003	0.0173	0.9937	0.0005	0.0216	0.9939	0.0003	0.0179
80% - 20%	0.9971	0.0001	0.0125	0.9972	0.0001	0.0129	0.9971	0.0001	0.0126
70% - 30%	0.9947	0.0003	0.0167	0.9933	0.0003	0.0185	0.9943	0.0003	0.0172
60% - 40%	0.9835	0.0008	0.0286	0.9834	0.0009	0.0316	0.9834	0.0009	0.0299

The results in Table 7 show that the proposed hybrid SARF model has improved the prediction accuracy and minimized the prediction error of the RF algorithm on all four data partitions. Comparing the results in Table 7 to the ones presented in Table 6, the proposed hybrid SARF model has an improved performance by recording up to 99% accuracy on first three data partitions (90% training - 10% testing, 80% training - 20% testing, and 70% training - 30% testing) and 98% accuracy on the fourth data partition (60% training - 40% testing). It can be observed that the proposed hybrid model has improved the prediction accuracy of the RF algorithm by 5.97%, 5.37%, 5.93%, and 5.73% on 90% - 10% data partition, 80% - 20% data partition, 70% - 30% data partition, and 60% - 40% data partition, respectively. To visualize the outputs of both the baseline and proposed SARF models,

the overall prediction performance of each model on 80% training and 20% testing data partition was compared to the actual data. Plots of these comparisons are presented in figures 6 and 7.

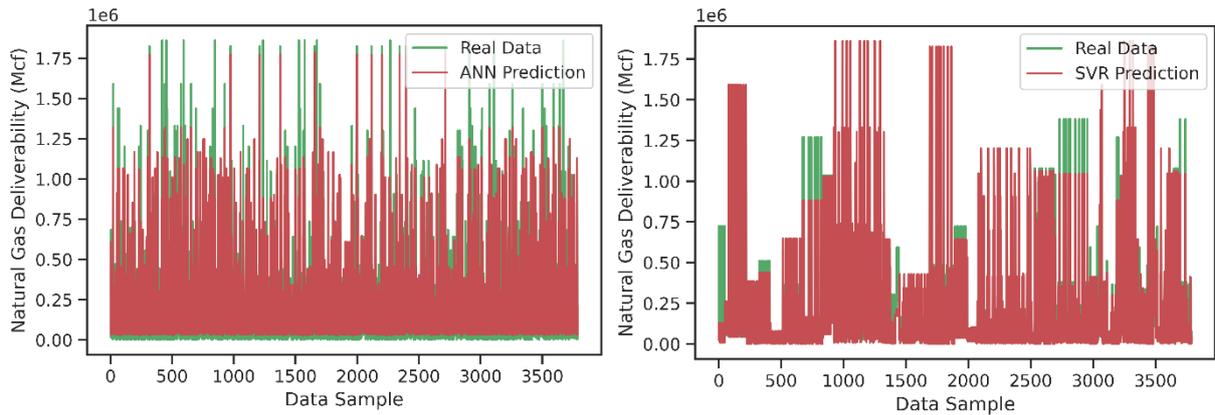


Fig. 6. Comparison of ANN prediction versus real data (left), and SVR prediction versus real data (right).

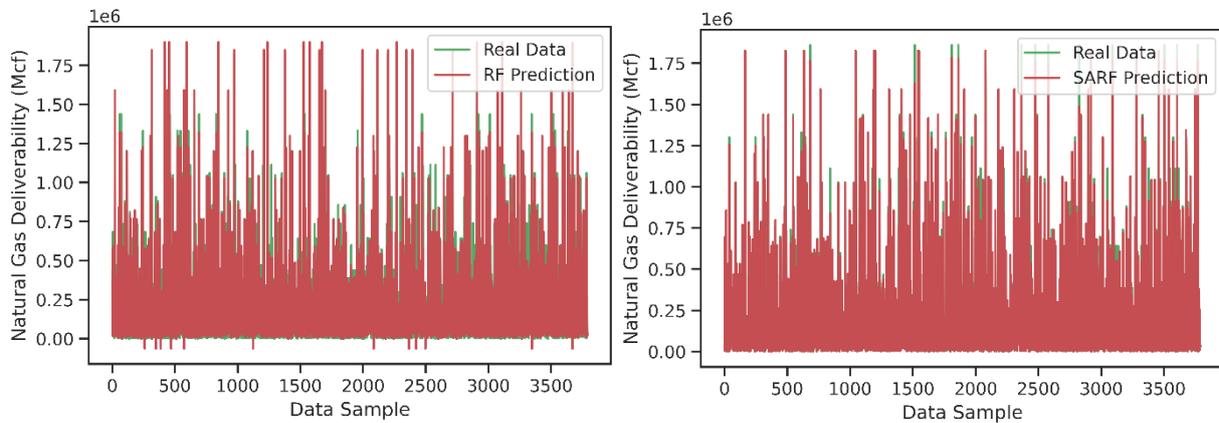


Fig. 7. Comparison of RF prediction versus real data (left), and SARF prediction versus real data (right).

From the results shown in Figure 6, it can be observed that the median deliverability values predicted by ANN model match well with the real data, however, there is a mismatch between the real data samples and the ones predicted by the ANN model for lower and higher deliverability values for the entire data samples. SVR model showed a good agreement between the real and predicted data for all lower deliverability values. However, there is a mismatch in some data samples with deliverability values of 0.5Mcf to 1.75 Mcf. From figure 7, a good agreement was observed for deliverability values predicted by RF, especially for samples with deliverability of 0.5 Mcf to 1.25 Mcf; however, there is a mismatch for some data samples with lower deliverability (0.5 Mcf to 0.00 Mcf) and higher deliverability (1.25 Mcf to 1.75 Mcf). On the other hand, the proposed SARF model showed a good agreement between the real and predicted data samples for all data samples with lower and middle deliverability, insignificant mismatch is observed for some data samples with higher values of deliverability. The developed SARF model benefits from the capabilities of two methods, SA and RF. The SARF model is a hybrid model that contains two units, one for learning and extracting significant features and the other for using the extracted feature to perform prediction. The advantage of the SARF hybrid model can be observed from the experimental results showed in Table 7, where the proposed SARF model achieved, an average 5.7% increase in accuracy on four separate data partitions over the baseline RF model. Furthermore, given that the SARF model maintains the dual effect of both SA and RF, it could encourage the widespread adoption of the SARF model for solving related problems.

5. Interpretation of SARF model using SHAP values

Research linked to XAI and interpretable ML has gained remarkable momentum over the last decade, however, very few examples have been applied in the energy sector [42], [43], [44], [45], and in particular within natural gas storage systems. As described in the results and discussion session, the SARF model demonstrated better performance and achieved higher accuracy on all four data partitions compared to the baseline models. However, it is challenging to explain the predictions of SARF out of the box. To mitigate this challenge, we apply the XAI technique known as SHAP to provide an intuitive explanation of the SARF model's inner functioning and increase its transparency. The explanations are based on the SARF model trained and tested on 80% training and 20% testing data partition. SHAP is a game theoretic technique for explaining the output of the ML model [46]. It provides global and local explanations using classic Shapley values from the game theory. In addition to feature importance estimate that focuses on interpreting a model in its entirety (global), the SHAP method provides interpretations of separate predictions of the whole model (local). Unlike the traditional feature importance methods that provide a traditional bar chart to explain AI/ML model performance, the SHAP method enables interpretation of even individual predictions of AI/ML model where a model's prediction is explained as a sum of values attributed to each feature. To demonstrate the importance and advantages of the SHAP method over traditional methods of explaining AI/ML model performance, the feature importance of the proposed SARF model for deliverability prediction is shown as a traditional bar chart in Figure 8 and as a SHAP summary plot in Figure 9. Even though both Figures 8 and 9 revealed that WGC has the highest value of feature importance, followed by BG and then TFC, it can be observed that Figure 9 (SHAP summary plot) reveals the SHAP values of every feature and for every sample (reservoir). This shows the importance of SHAP values in demonstrating a feature's responsibility for a change in the model output. The SHAP values are also helpful in generating dependence and individual force plots, as shown in Figures 10 and 11. These plots are vital for explaining how feature interaction impacts model output and why a particular sample (reservoir) receives its prediction, as explained in sections 5.2 and 5.3.

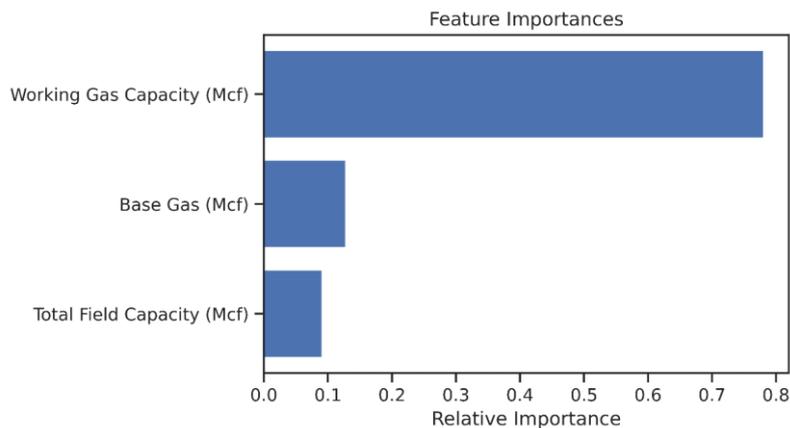


Fig. 8. The traditional feature importance plot for deliverability prediction by the proposed SARF model.

5.1 Global interpretation using SHAP summary plot

Here, SHAP values are combined to reveal the contribution of each predictor to the model's output (target variable). The SARF model's feature importance for predicting the deliverability of UGNS in depleted reservoirs is shown as a SHAP summary plot in Figure 9. This plot reveals the positive and negative impacts of the predictors on the target variable. The plot consists of dots, and each dot denotes a storage reservoir. The features are ranked in descending order of influence. The horizontal position reveals whether the impact of a feature is associated with a lower or higher prediction and the colours reveal whether that feature is low (in blue) or high (in red) for that reservoir. It can be inferred that WGC has the most significant impact on the model output, followed by BG, then TFC. This figure reveals that the higher the value of WGC for a particular storage reservoir, the larger the amount of natural gas to be delivered and vice versa. This information is essential for optimizing the deliverability of natural gas storage in the real world. It will enable reservoir engineers to focus data collection for subsequent modeling on features that are most impactful in maintaining optimal storage/withdrawal of natural gas.

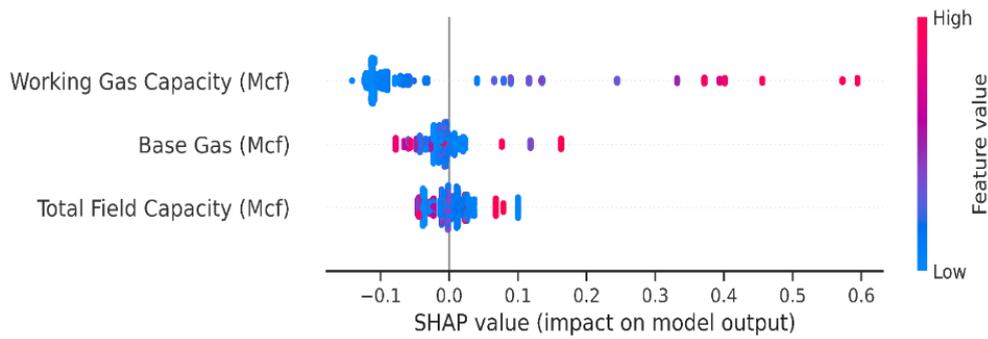


Fig. 9. SHAP summary plot showing feature influence on deliverability prediction. This figure shows a set of scatter plots for each feature, stacked by their order of importance. The y-axis refers to feature names in decreasing order of importance, and the x-axis indicates the SHAP values for each feature, ordered from lowest to highest. Each dot in the SHAP summary plot represents a sample in the data set, and its gradient color indicates the original value for that feature.

At the same time, it helps to avoid stock-outs or excess deliverability over time, narrowing the gap between supply and demand in the natural gas marketplace.

5.2. Feature interaction using SHAP dependence plot

It is also vital to show the marginal effect of one or two features on the predicted output; this can be achieved by creating a SHAP dependence plot for a particular feature. The plot can be used to interpret the effect of that particular reservoir feature on the reservoir's ability to deliver. Given that SHAP values are accountable for variation in the model output, the plot in Figure 10 shows the variation in predicted deliverability as WGC changes. It was observed that WGC has a relatively linear impact on model output. Furthermore, it was observed that the SHAP values for WGC were negative in a small number of reservoirs that have low WGC ($WGC < 0.15$), as seen from the group of blue dots in the bottom-left region of the plot. This trend can be interpreted as WGC having a negative impact on predicted deliverability in a small number of reservoirs with low WGC. In contrast, it had a positive impact on a large number of reservoirs with high WGC ($WGC \geq 0.15$). The broader spread of positive SHAP values for many reservoirs signifies the presence of feature interactions, as shown on the right Y-axis where WGC interacts with BG in this model. Furthermore, Figure 10 shows the effect of WGC on the predictions made by the model. Each dot is a single prediction (row) from the dataset. The x-axis shows the value of the WGC feature, and the y-axis shows its SHAP values. The SHAP values represent the impact a value of WGC has on the output of the model for a reservoir's prediction. The color corresponds to a second feature, BG, that has an interaction effect with WGC (by default, BG is chosen automatically by the SHAP algorithm) - suggesting an interaction effect between WGC and BG, as shown in a distinct vertical pattern of coloring. For the interaction effect shown in Figure 10, storage reservoirs with high values of WGC and BG are more likely to deliver a large amount of gas than those with low values of WGC and BG. This information is also essential for maintaining optimal storage/withdrawal of natural gas, while it helps to avoid stock-outs or excess deliverability over time.

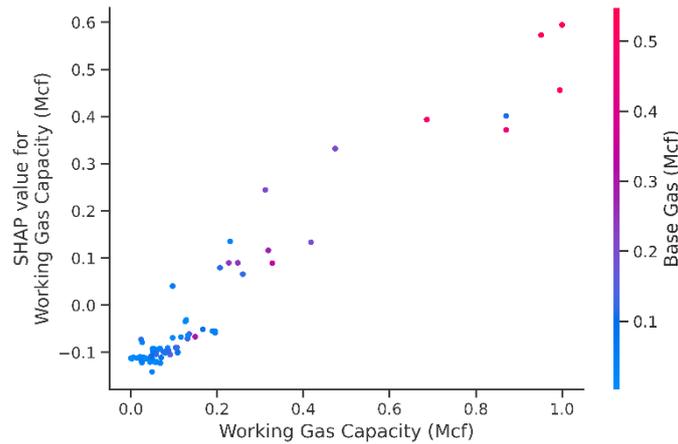


Fig. 10. SHAP dependence plot for WGC showing its impact on the model output and its interaction with BG. This figure presents a dependence plot between WGC on the X-axis and its SHAP values on the Y-axis. Each dot on the plot denotes a reservoir, and the gradient colours denote the original values of WGC from low (in blue) to high (in red).

5.3 Local interpretation using SHAP force plot

In Figure 10, the gradient colours that give a picture of the impact of reservoir feature on the model output might be a bit technical to be apprehended by an average reservoir production analyst. Hence, SHAP force charts were created for separate reservoirs to mitigate this shortcoming and understand how the model arrived at making a decision. SHAP force plot demonstrates how the input features influence the model's output. Thus, each reservoir gets its own set of SHAP values that explain why that reservoir receives its prediction/output and the predictors' contributions. Figure 11 shows two force plots for two separate reservoirs and their predicted deliverability values.

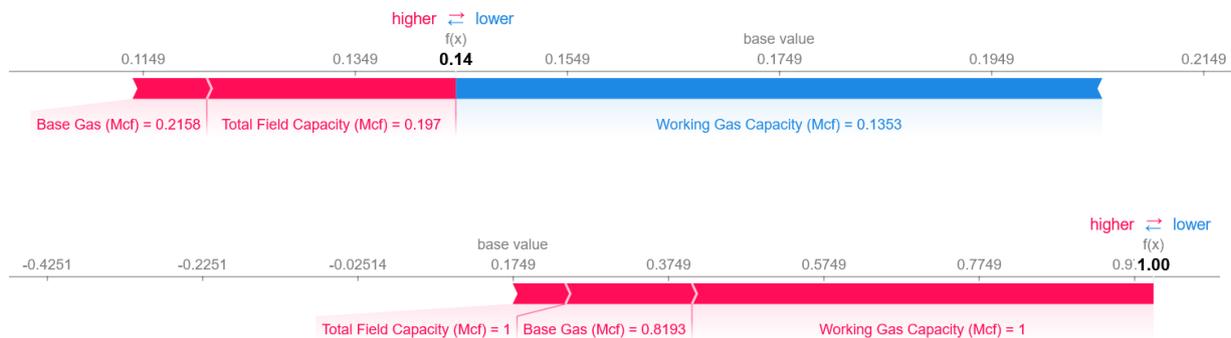


Fig. 11. Individual SHAP force plot to predict low (top) and high (bottom) deliverability values for UGNG storage in depleted reservoirs.

Each of the two plots in Figure 11 shows a range of reservoir deliverability values. The plots contain two values, one as a base value and the other as $f(x)$ (model prediction). The base value denotes the average model output over the training dataset, and the model prediction denotes the reservoir's deliverability predicted by the model. The base value can also be interpreted as the value that would be predicted if the reservoir features are not known. These visualisations showed the features responsible for the disparity between the predicted deliverability and the base value. Features that push the prediction lower are displayed in blue, and those that push the prediction higher are in red. These plots can provide recommendations that will inform decision-makers whether or not a particular storage reservoir will deliver a low or high volume of natural gas. For the storage reservoir shown in the first plot (top) of Figure 11, the model predicted a deliverability value of 0.14 Mcf, which is lower than the base value of 0.1749 Mcf. Hence, it indicates that the reservoir delivers a low volume of natural gas. This force plot also shows the features that cause the reservoir to deliver a low volume of natural gas. It can be observed from the plot that this reservoir has a BG = 0.2158 Mcf and TFC = 0.197 Mcf. These features push the reservoir to deliver a high volume of natural gas, as shown in red colours. Even with this force pushing towards higher volumes, a much

larger force pushes the reservoir to deliver low volumes, as shown in the blue colour bar. The forces that push the reservoir to deliver low or high volumes come from the SHAP values of each reservoir feature, denoted by the width of their respective bars. For this particular reservoir, it can be observed that it delivers a low volume of natural gas given that it has a WGC of 0.1357 Mcf, TFC of 0.197, and BG of 0.2158 Mcf. This interpretation agrees with what was observed in section 5.2, where results showed that WGC had a negative impact (pushing towards lower values) on predicted deliverability in a small number of reservoirs with low WGC ($WGC < 0.15$). Similarly, the model predicted a higher value of deliverability in the second plot (bottom) of Figure 11; it can be seen from the model's predicted value of 1.00 Mcf, which is greater than the base value. SHAP force plot can also be used to identify which feature attributes cause the reservoir to deliver a high volume of natural gas. The reservoir in the second plot (bottom) of Figure 11 has TFC of 1 Mcf, BG of 0.8193 Mcf, and WGC of 1 Mcf; its features push it to deliver a high volume of natural gas as they push the prediction towards higher values. Therefore, it also agrees with the discussion of section 5.2, where results showed that WGC had a positive impact (pushing towards higher values) on predicted deliverability in a large number of reservoirs with high WGC ($WGC \geq 0.15$). Thus, it can be said that the reservoir in the second plot (bottom) of Figure 11 belongs to those reservoirs that are impacted positively by WGC, given its value of WGC which is 1 Mcf.

6. Conclusion

This paper reports the capability of a hybrid data-driven modelling technique referred to as SARF that demonstrated an error reduction and accuracy improvement compared to RF, ANN, and SVR techniques. It was found that the proposed SARF model achieved an average 5.7% increase in accuracy on four separate data partitions in predicting natural gas deliverability in UNGS. Thereby results in a more accurate model, with less discrepancy between actual and predicted values of deliverability. The global feature importance, feature interaction, and local interpretation analysis, together with associated visualisations presented using the SHAP method, provided an insight into the inner workings of the prediction model used in this work. The proposed SARF model was implemented on 3744 actual data samples from 52 active storage facilities in the different States of the US. Bulk energy storage facilities such as depleted reservoirs play a vital role in meeting seasonal energy demands. As underground energy storage facilities are essential for balancing out supply and demand, the outcome of this study will benefit the natural gas industry when making decisions with respect to optimal reserve maintenance, transport cost minimization, and adjustment in supply and demand. Also, this study will stimulate further research in exploring the applicability of other ML algorithms in bulk energy storage facilities and serve as a reference to ML and energy storage researchers. Future work will build on this by exploring the applicability of advanced ML algorithms in predicting the deliverability of other energy storage facilities.

Acknowledgment

The authors acknowledge financial support from the Petroleum Technology Development Fund (PTDF) of the Federal Republic of Nigeria. Special thanks to the subject Editor and anonymous reviewers for their valuable time and insightful comments, which significantly improved the quality of this paper.

Data Availability

Datasets related to this article can be found at

<https://www.eia.gov/naturalgas/ngqs/#?report=RP8&year1=2016&year2=2021&company=Name>, an official repository of the U.S. Energy Information Administration (EIA).

References

- [1] IEA (2022), Gas Market Report Q1-2022: Including Gas Market Highlights 2021, OECD Publishing, Paris, <https://doi.org/10.1787/4a6dc268-en>.
- [2] H. Plaat. Underground gas storage: Why and how. *Geol. Soc. Spec. Publ.*, vol. 313, pp. 25–37, 2009, doi: 10.1144/SP313.4.
- [3] O'Shea, Richard & Lin, Richen & Wall, David M. & Browne, James D. & Murphy, Jerry D. Using biogas to reduce natural gas consumption and greenhouse gas emissions at a large distillery. *Appl. Energy*, vol. 279, 2020, doi: 10.1016/j.apenergy.2020.115812.
- [4] M. Healey. Underground Gas Storage in Cheshire – the Costain Experience. November, 2008.
- [5] D. J. Evans and R. A. Chadwick. Underground gas storage: An introduction and UK perspective. *Geol. Soc. Spec. Publ.*, vol. 313, no. 1, pp. 1–11, 2009, doi: 10.1144/SP313.1.

- [6] U.S. Energy Information Administration. The Basics of Underground Natural Gas Storage - U.S. Energy Information Administration. *Novemb.* 16 , pp. 1–8, 2015.
- [7] F. Verga. What’s conventional and what’s special in a reservoir study for underground gas storage. *Energies*, vol. 11, no. 5, 2018, doi: 10.3390/en11051245
- [8] A. Ali, U. Diala and L. Guo. Data-Driven Based Modelling of Pressure Dynamics in Multiphase Reservoir Model. *2022 UKACC 13th International Conference on Control (CONTROL)*, 2022, pp. 189-194, doi: 10.1109/Control55989.2022.9781447.
- [9] M. Davidson. Underground gas storage project at Welton oilfield, Lincolnshire: Local perspectives and responses to planning, environmental and community safety issues. *Geol. Soc. Spec. Publ.*, vol. 313, pp. 149–161, 2009, doi: 10.1144/SP313.10.
- [10] D. J. Evans. A review of underground fuel storage events and putting risk into perspective with other areas of the energy supply chain,” *Geol. Soc. Spec. Publ.*, vol. 313, pp. 173–216, 2009, doi: 10.1144/SP313.12.
- [11] S. Diampovesa, A. Hubert, and P. A. Yvars. Designing physical systems through a model-based synthesis approach. Example of a Li-ion battery for electrical vehicles. *Comput. Ind.*, vol. 129, p. 103440, 2021, doi: 10.1016/j.compind.2021.103440.
- [12] A. Ali and L. Guo. Adaptive neuro-fuzzy approach for prediction of dewpoint pressure for gas condensate reservoirs. *Pet. Sci. Technol.*, vol. 38, no. 9, pp. 673–681, 2020, doi: 10.1080/10916466.2020.1769655.
- [13] F. Chiarello, P. Belingheri, and G. Fantoni. Data science for engineering design: State of the art and future directions. *Comput. Ind.*, vol. 129, p. 103447, 2021, doi: 10.1016/j.compind.2021.103447.
- [14] A. Ali, M. K. Ahmed, K. Aliyuda and A. M. Bello. Deep Neural Network Model for Improving Price Prediction of Natural Gas. *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*, 2021, pp. 113-117, doi: 10.1109/ICDABI53623.2021.9655885.
- [15] A. Ali. Ensemble Learning Model for Prediction of Natural Gas Spot Price Based on Least Squares Boosting Algorithm. *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, 2020, pp. 1-6, doi: 10.1109/ICDABI51230.2020.9325615.
- [16] X. Zhou, W. Lin, R. Kumar, P. Cui, and Z. Ma. A data-driven strategy using long short term memory models and reinforcement learning to predict building electricity consumption. *Appl. Energy*, vol. 306, no. PB, p. 118078, 2022, doi: 10.1016/j.apenergy.2021.118078.
- [17] W. Yang, J. Shi, S. Li, Z. Song, Z. Zhang, and Z. Chen. A combined deep learning load forecasting model of single household resident user considering multi-time scale electricity consumption behavior. *Appl. Energy*, vol. 307, no. November 2021, p. 118197, 2022, doi: 10.1016/j.apenergy.2021.118197.
- [18] Q. Chen, Z. Kuang, X. Liu, and T. Zhang. Energy storage to solve the diurnal , weekly , and seasonal mismatch and achieve zero-carbon electricity consumption in buildings. *Appl. Energy*, vol. 312, no. November 2021, p. 118744, 2022, doi: 10.1016/j.apenergy.2022.118744.
- [19] B. Svetozarevic, C. Baumann, S. Muntwiler, L. Di Natale, M. N. Zeilinger, and P. Heer. Data-driven control of room temperature and bidirectional EV charging using deep reinforcement learning : Simulations and experiments. *Appl. Energy*, vol. 307, no. November 2021, p. 118127, 2022, doi: 10.1016/j.apenergy.2021.118127.
- [20] Y. Hu, X. Cheng, S. Wang, J. Chen, T. Zhao, and E. Dai. Times series forecasting for urban building energy consumption based on graph convolutional network. *Appl. Energy*, vol. 307, no. November 2021, p. 118231, 2022, doi: 10.1016/j.apenergy.2021.118231.
- [21] G. I. Tsoumalis, Z. N. Bampos, P. N. Biskas, S. D. Keranidis, P. A. Symeonidis, and D. K. Voulgarakis. A novel system for providing explicit demand response from domestic natural gas boilers. *Appl. Energy*, vol. 317, no. March, p. 119038, 2022, doi: 10.1016/j.apenergy.2022.119038.
- [22] M. Kora. Machine learning analysis of electric arc furnace process for the evaluation of energy efficiency parameters. *Appl. Energy*, vol. 307, no. December 2021, 2022, doi: 10.1016/j.apenergy.2021.118209.
- [23] M. Al-Qatf, Y. Lasheng, M. Al-Habib, and K. Al-Sabahi. Deep Learning Approach Combining Sparse Autoencoder with SVM for Network Intrusion Detection. *IEEE Access*, vol. 6, pp. 52843–52856, 2018,

doi: 10.1109/ACCESS.2018.2869577.

- [24] I. D. Mienye, Y. Sun, and Z. Wang. Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. *Informatics Med. Unlocked*, vol. 18, p. 100307, 2020, doi: 10.1016/j.imu.2020.100307.
- [25] H. Sewani and R. Kashef. An Autoencoder-Based Deep Learning Classifier for Efficient Diagnosis of Autism. *Children*, vol. 7, no. 10, p. 182, 2020, doi: 10.3390/children7100182.
- [26] H. Liu and R. Yang. A spatial multi-resolution multi-objective data-driven ensemble model for multi-step air quality index forecasting based on real-time decomposition. *Comput. Ind.*, vol. 125, p. 103387, 2021, doi: 10.1016/j.compind.2020.103387.
- [27] B. Scheibel, J. Mangler, and S. Rinderle-Ma. Extraction of dimension requirements from engineering drawings for supporting quality control in production processes. *Comput. Ind.*, vol. 129, p. 103442, 2021, doi: 10.1016/j.compind.2021.103442.
- [28] P. Calvo-Bascones, M. A. Sanz-Bobi, and T. M. Welte. Anomaly detection method based on the deep knowledge behind behavior patterns in industrial components. Application to a hydropower plant. *Comput. Ind.*, vol. 125, p. 103376, 2021, doi: 10.1016/j.compind.2020.103376.
- [29] H. Wu. A Deep Learning-Based Hybrid Feature Selection Approach for Cancer Diagnosis. *J. Phys. Conf. Ser.*, vol. 1848, no. 1, p. 012019, 2021, doi: 10.1088/1742-6596/1848/1/012019.
- [30] A. Ali. Data-driven based machine learning models for predicting the deliverability of underground natural gas storage in salt caverns. *Energy*, vol. 229, p. 120648, 2021, doi: 10.1016/j.energy.2021.120648.
- [31] [30] Y. Onal. Gaussian Kernel Based SVR Model for Short-Term Photovoltaic MPP Power Prediction. *Comp. Systems Sci. & Eng.*, vol. 41, No.1, pp. 141-156, 2022, doi: 10.32604/csse.2022.020367.
- [32] X. Huang, K. Sun, and X. Xiao. A Neural Network-Based Power Control Method for Direct-Drive Wave Energy Converters in Irregular Waves. *IEEE Trans. Sustain. Energy*, vol. 11, no. 4, pp. 2962–2971, 2020, doi: 10.1109/tste.2020.2984328.
- [33] Breiman, L. Random Forests. *Machine Learning*, 45, 5-32, 2021, <http://dx.doi.org/10.1023/A:1010933404324>
- [34] A. C. Cameron and F. A. G. Windmeijer. An R-squared measure of goodness of fit for some common nonlinear regression models. vol. 00, no. i, pp. 329–342, 1996.
- [35] C. Voyant *et al.* Machine learning methods for solar radiation forecasting : A review. *Renew. Energy*, vol. 105, pp. 569–582, 2017, doi: 10.1016/j.renene.2016.12.095.
- [36] U.S. Energy Information Administration (EIA). <https://www.eia.gov/naturalgas/ngqs/#?report=RP8&year1=2016&year2=2021&company=Name> (Retrieved on 02/01/2022).
- [37] T. M. Letcher, *Storing Energy: With Special Reference to Renewable Energy Sources*. 2016.
- [38] D. J. Evans. An appraisal of Underground Gas Storage technologies and incidents , for the development of risk assessment methodology. *Br. Geol. Surv. Open Rep. OR/07/023*, p. 287, 2007, doi: 10.1016/j.jfca.2010.03.028.
- [39] K. Aliyuda, J. Howell, A. Hartley, and A. Ali. Stratigraphic controls on hydrocarbon recovery in clastic reservoirs of the Norwegian Continental Shelf. *Pet. Geosci.*, pp. petgeo2019-133, 2020, doi: 10.1144/petgeo2019-133.
- [40] K. Aliyuda, J. Howell, and E. Humphrey. Impact of Geological Variables in Controlling Oil-Reservoir Performance: An Insight from a Machine-Learning Technique. *SPE Reserv. Eval. Eng.*, no. November 2019, pp. 1–14, 2020, doi: 10.2118/201196-pa.
- [41] E. L. Rawlins and M. A. Schellhardt. Back-pressure data on natural-gas wells and their application to production practices. 1935.
- [42] P. Arjunan, K. Poolla, and C. Miller. EnergyStar ++ : Towards more accurate and explanatory building energy benchmarking. *Appl. Energy*, vol. 276, no. January, p. 115413, 2020, doi:

- 10.1016/j.apenergy.2020.115413.
- [43] S. Wenninger, C. Kaymakci, and C. Wiethe. Explainable long-term building energy consumption prediction using QLattice. *Appl. Energy*, vol. 308, no. December 2021, p. 118300, 2022, doi: 10.1016/j.apenergy.2021.118300.
 - [44] C. Miller. Energy & Buildings What ' s in the box ?! Towards explainable machine learning applied to non-residential building smart meter classification. *Energy Build.*, vol. 199, pp. 523–536, 2019, doi: 10.1016/j.enbuild.2019.07.019.
 - [45] G. Mitrentsis and H. Lens. An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting. *Appl. Energy*, vol. 309, no. December 2021, p. 118473, 2022, doi: 10.1016/j.apenergy.2021.118473.
 - [46] Lundberg, S. M. et al. Explainable AI for trees: From local explanations to global understanding. arXiv preprint arXiv:1905.04610 (2019).