Application of machine learning algorithms in classification the flow units of the Kazhdumi reservoir in one of the oil fields in southwest of Iran

Fatemeh mohammadinia[1,2] · Ali Ranjbar[1,2] · Moein Kafi[1,2] · Reza Keshavarz[1]

Abstract

By determining the hydraulic flow units (HFUs) in the reservoir rock and examining the distribution of porosity and permeability variables, it is possible to identify areas with suitable reservoir quality. In conventional methods, HFUs are determined using core data. This is while considering the non-continuity of the core data along the well, there is a great uncertainty in generalizing their results to the entire depth of the reservoir. Therefore, using related wireline logs as continuous data and using artificial intelligence methods can be an acceptable alternative. In this study, first, the number of HFUs was determined using conventional methods including Winland R35, flow zone index, discrete rock type and k-means. After that, by using petrophysical logs and using machine learning algorithms including support vector machine (SVM), artificial neural network (ANN), LogitBoost (LB), random forest (RF), and logistic regression (LR), HFUs have been determined. The innovation of this article is the use of different intelligent methods in determining the HFUs and comparing these methods with each other in such a way that instead of using only two parameters of porosity and permeability, different data obtained from wireline logging are used. This increases the accuracy and speed of reaching the solution and is the main application of the methodology introduced in this study. Mentioned algorithms are compared with accuracy, and the results show that SVM, ANN, RF, LB, and LR with 90.46%, 88.12%, 91.87%, 94.84%, and 91.56% accuracy classified the HFUs respectively.

Keywords Hydraulic flow units · Flow Zone Index · Rock type determination · Machine learning · Artificial intelligence

List of symbols

b	Bias value
C_j	Center of clusters
DRT	Discrete Rock Type
FZI	Flow Zone Index
F(Output$_i$)	The value of the sigmoid function
h_i	Decision tree model
J	Objective function
K	Permeability
Output$_i$	Output layers
P	Similar to the density distribution function
RQI	Reservoir Quality Index
W	Weight
X	Input data
x	Combination of the classification model
Y	Output variable
Z	Integer
β_i	Regression coefficients
β_0	Members that represent the reference level
φ_e	Effective porosity
ε_i and ε_i^*	Ineffective variables
π	Probability of an event

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
ANFIS	Artificial Network Fuzzy Inference System
BVW	Bulk Volume Water
DRT	Discrete Rock Type
EOR	Enhanced Oil Recovery
FZI	Flow Zone Index
HFU	Hydraulic Flow Unit
IMLR	Iterative Multi-Linear Regression
LB	LogitBoost
LR	Logistic Regression

✉ Ali Ranjbar
ali.ranjbar@pgu.ac.ir

1   Department of Petroleum Engineering, Faculty of Petroleum, Gas and Petrochemical Engineering, Persian Gulf University, Bushehr, Iran

2   Sustainable Exploitation of Underground Resources Research Core, Persian Gulf University, Bushehr, Iran

MICP   Mercury Injection Capillary Pressure
MLP   Modified Lorenz Plot
MRGC   Multi-Resolution Graph-Based Clustering
ML   Machine Learning
NPI   Normalized Porosity Index
PMR   Pore to Matrix Ratio
SSE   Sum of Squares Error
SVM   Support Vector Machine
SMLP   Stratigraphic Modified Lorenz Plot
SFP   Stratigraphic Flow Profile
RF   Random Forest
RQI   Reservoir Quality Index
RPI   Reservoir Potential Index
RFN   Rock Fabric Number

Introduction

As hydrocarbon reservoirs have heterogeneity in macroscopic and microscopic scales, an accurate description requires a full study of reservoir uniformity. One of the most popular techniques used to describe the characteristics and reservoir heterogeneity is to determine the number of flow units. In drilling, production and reservoir studies, determination of reservoir rock types, and the number of flow units are very important. Hydraulic flow units are part of a reservoir with unique characteristics that have a significant role in fluid flow in reservoir. They may be interacting with other flow units. The reservoir quality and the rock type are determined by this feature and the relationship between porosity and permeability. More accurate estimation of porosity and permeability requires correct classification of flow units. (Ebanks Jr 1987, Amaefule et al. 1993, Guo et al. 2005, Tiab and Donaldson 2015; Elnaggar 2018, Sharifi-Yazdi et al. 2020).

Flow units are mainly used to describe the hydrocarbon reservoir. Flow units determination is necessary for accurate reservoir petrophysical modeling (Hosseini Bidgoli et al. 2014). The rock types are determined to a reservoir classification into separate units that have deposited with similar diagenetic changes or have the same geological status. To have a more accurate estimate of the relevance between permeability and porosity and more realistic simulation results, the flow units must be determined correctly (Guo et al. 2005, Zargari et al. 2013).

In order to determine hydraulic flow units (HFUs) of reservoirs, Amaefule et al. introduced a new technique using the Kozeny-Carman equation and the medium hydraulic radius. This technique creates a line with a fixed slope for each hydraulic unit in the Reservoir Quality Index (RQI) versus Pore to Matrix Ratio (PMR) diagram. The intersection point of the line created by the Normalized Porosity Index (NPI) with line PMR = 1 is called the FZI. This parameter indicates a unique HFU. After that the correlation of permeability and FZI could by calculated by regression models (Amaefule et al. 1993). Using RQI and FZI, Gunter et al. show that the rock type is extremely beneficial in permeability and initial water saturation modeling which are used in geological modeling and reservoir simulations. They introduced acceptable graphical methods and used them to identify the rock type and analyze the flow unit in carbonate and sandstone reservoirs (Gunter et al. 1997). However, although the method used in that study led to acceptable results, determining the rock types in sandstone reservoirs still requires a more comprehensive method.

Abnavi et al. identified the number of gas reservoir flow units in southern Iran using histogram analysis methods and normal probability diagrams. The result of the study shows that the normal probability diagram is a more reliable method for detecting HFUs (Abnavi et al. 2018). Shalaby et al. study, using core data (porosity and permeability) of Qasr field, various methods were used to analyze and describe the sandstone of the Khatatba Formation. In their study, the number of flow units has been defined by the RQI, FZI, and NPI. The Winland R35 equation was used to describe the geometry of the pores and the diameter of the pore throat, which eventually led to the classification of sandstones into three classes of flow units and three different rock types (Shalaby 2021). Moghtadar et al. used the concept of hydraulic flow units and electric current units to describe and evaluate the sandstone reservoir of the Nubia Formation in Gebel Abu Hassle. They also determined the number of flow units and the rock type using RQI and Winland R35 methods (El-Sayed et al. 2021). Nayak et al. collected porosity and permeability data from 32 core samples of the calcareous field from four different Mumbai regions. The porosity range of these samples was from 0.3% to 20.5%, and the permeability range was from 0.002 to 1.484 millidarcy, and the depth was 1618.86 to 1634.14 m. Using the obtained data, the FZI is calculated, and the HFUs are determined. The least squares regression (LSR) method has been used to determine the flow units (Nayak et al. 2021).

In the meantime, many researchers have also used machine learning knowledge in their studies. In 2020, Khadem et al. developed a system for detecting the rock type and the HFUs of detrital reservoirs with uniform pores. The system is being implemented on an oil field in the Persian Gulf. First, physical models, classify field rocks into three types with different characteristics. Then, using the core data, the number of flow units was calculated and expanded the information obtained by using simultaneous inversion and the rock physics models throughout the reservoir (Khadem et al. 2020). Sengel et al. developed a dynamic model to predict the future performance of the Germik reservoir in southeastern Turkey. At the first, the hydraulic flow units instead of the reservoir facies model were determined in cored wells,

and then using artificial neural networks (ANN), the flow units for the other wells and through the model were estimated. The results are used to build a reservoir permeability model. Finally, the results show that the date-simulated model can be safely used for enhanced oil recovery (EOR) screening (Sengel and Turkarslan 2020). In 2021, Abnavi et al. determined the number of flow units in the hydrocarbon field in the south of Iran using the core data by the FZI method. Then, using an artificial network fuzzy inference system (ANFIS), permeability of the studied well was estimated. ANFIS estimates the permeability with an error of 1.83%. This algorithm estimates the permeability of noncored wells with a 21.5% error (Abnavi 2021). A summary of some of the articles studied is given in Table 1.

Based on studies of the literature and previous studies in this field, the importance of studying flow units is determined. Such cases are very important in the studies of enhanced oil recovery. In order to study more about the

issue, it is referred to the studies of Wu et al. (2018) and Wu et al. (2016). It is obvious that in most studies, conventional methods have been used to determine the rock types, and studies on new methods of machine learning (ML) in this field need more research. Conventional methods for classifying the number of flow units require direct core test data such as porosity and permeability. This is even though coring, operations in reservoir formations are carried out in a limited number of field wells due to their high cost and time-consuming nature, and it is not possible to access the cores of different parts of a reservoir in the oil field. This causes the conventional methods of classifying flow units to work with fewer input features and so not provide accurate and acceptable results. As an alternative, machine learning methods can be used for this purpose since they use well log data besides core data (given that petrophysical log data are available in most wells and contain information from well columns).

Table 1 Summary of some of related articles studied in HFU determination

No	Input data	Used methods	Output	Year (refs)
1	- Porosity - Permeability	- FZI	- HFUs - Estimated permeability	Amaefule et al. (1993)
2	- Porosity - Pore size distribution - Water saturation	- Iterative multi-linear regression (IMLR)	- HFUs - Estimated permeability	Al-Ajmi and Holditch (2000)
3	- Petrophysical logs - Production data	- DRT - LLNFM	- HFUs - Rock types	Ghadami et al. (2015)
4	- Porosity - Permeability	- ANN - MRGC	- HFUs - Estimated permeability	Nouri-Taleghani et al. (2015)
5	- Porosity - Permeability	- FZI - MICP	- HFUs - Rock types	Nabawy et al. (2018)
6	- Porosity - Permeability	- Histogram analysis - Normal probability plot	- HFUs	Abnavi et al. (2018)
7	- Porosity - Permeability	- Rock fabric number (RFN) - Winland R35—FZI - BVW -SMLP - MLP—SFP	- HFUs - Location of HFUs - Quality of the reservoir - Reservoir heterogeneity analysis	Riazi (2018)
8	- Permeability - Pressure, - Oil water contact	- ANN - FZI	- HFUs - Estimated permeability	Sengel and Turkarslan (2020)
9	- Petrophysical logs - Seismic data - Porosity - Permeability	- Inversion	- HFUs	Khadem et al. (2020)
10	- Porosity - Permeability	ANFIS, FZI	- HFUs - Estimated permeability	Abnavi (2021)
11	- Porosity - Permeability	- Winland R35	- HFUs - Quality of the reservoir	Shalaby (2021)
12	- Porosity - Permeability	- Winland R35	- HFUs and EFUs - Quality of the reservoir	El-Sayed et al. (2021)
13	- Porosity - Permeability	- FZI, - LSR	- HFUs	Nayak et al. (2021)

In this study, HFUs have been classified by using well log data and machine learning methods. If most of the previous studies have used core data for this issue, considering that there is no core data for the entire length of the well and access to core data is very costly and time-consuming, replacing petrophysical log data with core data is a suitable method for classifying HFUs. For this purpose, the well log and core data have been collected. Then, the number of HFUs was determined using the conventional methods of Winland R35, FZI, DRT, and k-means. Machine learning methods including ANN, support vector machine (SVM), LogitBoost, logistic regression, and random forest (RF) have been studied and used to classify HFUs considering the classification calculated from the FZI method as the optimal classification of HFU reservoir. Finally, the performance of these methods has been compared. According to the machine learnings' performance in the classification of flow units, this method can be extended to the entire length of the well and the flow units of points of the well that lack core data can be predicted. Among the innovations of this research is the use of various applied machine learning methods in the classification of HFUs and the comparison of these methods and their performance in the Kazhdumi Formation, which has sandy shale facies (only in some southwestern fields of Iran).

Case study and available data

The studied field is located in the coastal part of the Persian Gulf sedimentary basin called the Khark Romeyle. The Persian Gulf is an epi-continental and marginal sedimentary basin that is in multiple sedimentary environments (Siebold 1969). The Persian Gulf is part of the Arabian plate, at the intersection of the Arabic and Eurasian lithosphere plates. The time of its formation in the current situation is the Late Miocene and dates back to the formation of the Zagros Mountains. The tectonics of this basin is similar to the tectonic conditions of the foreland basin on the edge of the Zagros Mountains. The deepest part of the Persian Gulf basin, from the Middle Jurassic to the Lower Cretaceous, is located the northwestern corner of the Persian Gulf (Rabbani 2013). The Persian Gulf basin can be introduced as one of the richest hydrocarbon basins in the world since more than 50% of the world's gas and oil are located in the Persian Gulf basin (Rabbani 2007). The studies field is an anticline trap with an almost NS trend which is located in NW of the Persian Gulf (Fig. 1). The reservoir formation of this field is Kazhdumi which is deposited in Early Albian to Middle Albian. Although this formation is often known as the source rock in the Zagros sedimentary basin with lithology of shale, the middle parts of this formation in the NW of the Persian Gulf include sandy sequences which could act as a high-potential reservoir rock (Motiei 1995). The Kazhdami reservoir in the Khark Romeyle basin is deposited in a shallow marine and deltaic environment. The lithology of this formation is composed of fine to coarse sandstone and has a highly faulted reservoir (Nairn and Alsharhan 1997).

The stratigraphic column of the NW of the Persian Gulf basin is shown in Fig. 2. In the studied area, from surface to depth, formations are Bakhtiari, Aghajari, Gachsaran, Asmari, Jahrom, Tarbour and Gurpi, Sarvak, Kozhdami, Darian Gadvan, Fahlian, and Heath Anhydrite, respectively.

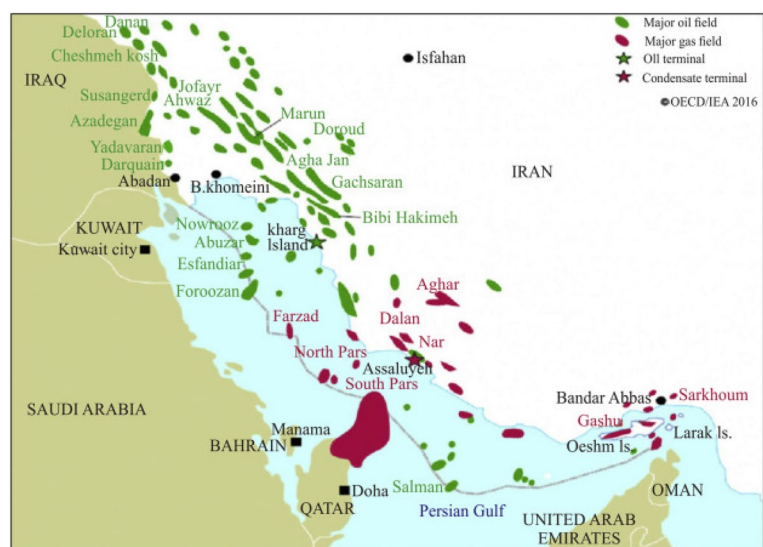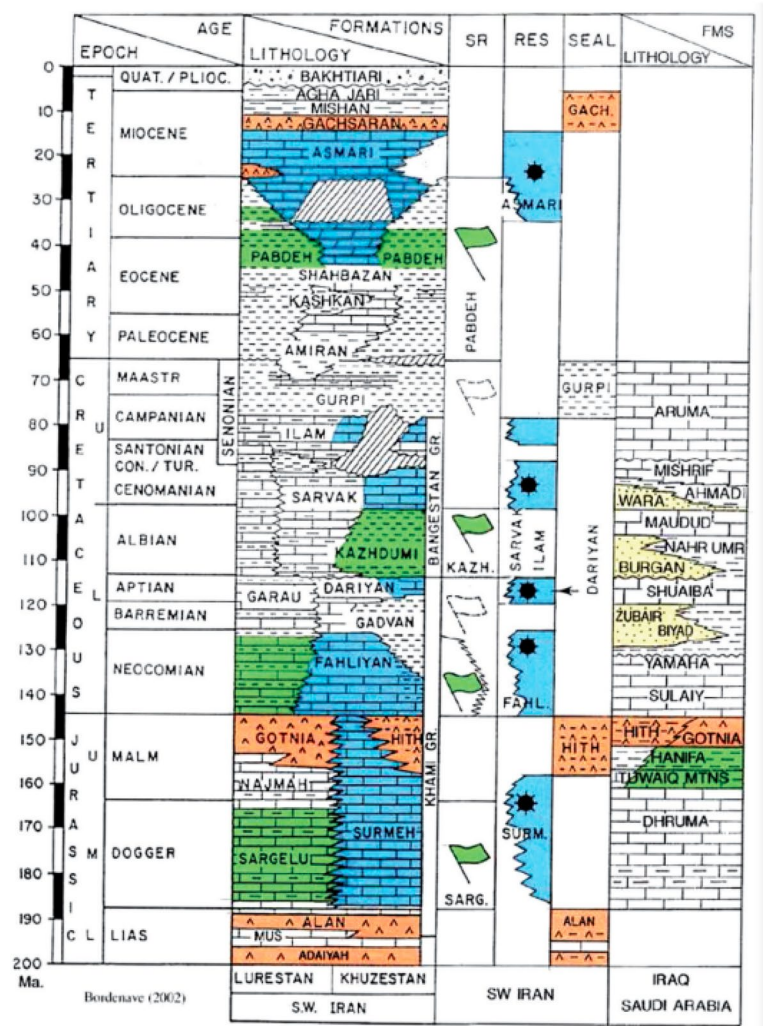**Fig. 1** Location of the southwestern Iran oil and gas fields (Zargar et al. 2020)

Fig. 2 Stratigraphic column of the studied area (Mohebian et al. 2013)

In this study, in order to classification of rock type, 212 core sample data including porosity and permeability and petrophysical well logs of an oil field in southwest if Iran have been used. The available logs were RT, DT, HCAL, NPHI, RHOZ, and PEFZ. The range of each input parameters such as porosity, permeability, and depth is reported in Table 2.

Table 2 Value range of available data

Data	Unit	Max	Min	Mean
Depth	M	2393.2	2157	2309.091
Porosity	%	34.1	2.1	20.94366
Permeability	mD	9944.3	0.1	1298.154
RT	Ohm.m	19.63894	0.50193	4.103192
CGR	API.GR	46.18848	4.79701	27.24397
DT	Us	120.627	57.58504	93.93078
HCAL	In	17.506	12.15448	13.36917
NPHI	Fraction	0.84436	0.11115	0.449366
RHOZ	gm/cc	2.79749	1.53385	2.320674
PEFZ	Barns/electron	9.77355	2.76304	5.624213

Methodology

Conventional methods

FZI method

As mentioned, reservoir rocks can be divided into several different flow units from a geological or engineering point of view to describe how they behave during different production applications (Gomes et al. 2008). Since the FZI depends on the geological properties and the geometry of the different rock types and is also a function of the reservoir quality and porosity ratio, it is a desirable parameter to determine hydraulic flow units (HFU) (Abed 2014). Each HFU has a specific value of FZI which is determined by log analysis (porosity and permeability

logs) and calculated from the RQI and normalized porosity (Yi et al. 2021):

$$RQI = 0.0314\sqrt{\frac{K}{\varphi_e}} \tag{1}$$

where K and φ_e are the permeability and effective porosity of the rock, respectively. The normalized porosity is obtained from Eq. 2 which is used in the FZI calculations (Amaefule, Altunbay et al. 1993):

$$\varphi_z = \frac{\varphi_e}{1 - \varphi_e} \tag{2}$$

Finally, the FZI is obtained by Eq. 3:

$$FZI = \frac{RQI}{\varphi_Z} \tag{3}$$

By applying mathematical operations, Eq. 4 can be deduced:

$$\log(RQI) = \log(\varphi_Z) + \log(FZI) \tag{4}$$

In the log(RQI) vs. $\log(\varphi_Z)$ diagram, all the FZl samples with similar values are placed on a straight line with the same slope. The points placed on a straight line have similar pore properties (Fig. 3). The constant FZI value could be obtained from the intersection point of the unit slope with φ_Z=1(Amaefule et al. 1993). To identify all the distributions presented in the original data, it is necessary to create a histogram of log(FZI). Since the FZI is multiple of all logarithmic normal distributions, the log(FZI) histogram represents n number of normal distributions for n-flow units. In situations where the clusters are distinctly separated, the

histogram can intelligibly identify apiece HFUs (Al-Ajmi and Holditch 2000; Abed 2014).

The normal probability diagram is used to evaluate the compliance of a set of data with a standard bell-shaped curve. To calculate the normal probability graph, log(FZI) data should be sorted. After that, percentiles with uniform distances from the normal distribution could be determined (Nouri-Taleghani et al. 2015). Since the FZI mean values cannot be reached from the probability plot, the FZI instance value of each HFU is obtained by averaging all the FZI values in the corresponding HFU range. It should be noted that the overlap effect may vary or deform straight lines in the probability plot (Al-Ajmi and Holditch 2000).

Winland R35 method

Winland defined his equation using 300 samples from the Spindle field. In 1972, by examining different mercury saturations, he showed that the best value for mercury saturation is 35% for calculating the pores radius which show the best path for fluid flow (Winland 1972). The performance of this method is based on capillary pressure curves (Soleymanzadeh et al. 2019).

Winland calculated the most appropriate curve at 35% mercury saturation by regression analysis to establish an equation between porosity, permeability and the size of the pore throat which leaded to Eq. 5 (Winland 1972; Kolodzie 1980):

$$\log(R35) = 0.735 + 0.588 log(k) - 0.864 log(\varphi) \tag{5}$$

By this equation, the data can be categorized and the quality of the reservoir determined based on the size of the pore throats (Spearing et al. 2001).

DRT method

Using the Winland equation, the continuous values of FZI are converted to discrete ones. Following the discretization of the FZI values, using Eq. 6, the core data are classified into separate categories. The equation mentioned by Chakani and Kharat in 2012 was used for carbonate reservoirs (Chekani and Kharrat 2012):

$$DRT = Round(2\log(FZI) + 10.7) \tag{6}$$

It should be noted that this equation is also used to predict permeability in reservoir static modeling. FZI values are determined in the reservoir grid blocks, and the obtained DRT values from Eq. 6 are propagated through the model. According to the relation between porosity and permeability in each DRT set, a certain amount of permeability is assigned to each reservoir grid blocks (Chekani and Kharrat 2012).
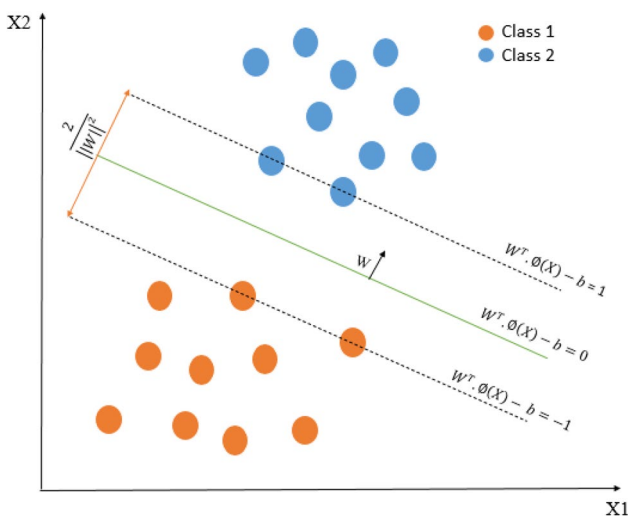

Fig. 3 Support vector machine hyper-plane

K-means method

K-means is an unsupervised algorithm that can easily divide a data set into several separate subsets (MacQueen 1967). This method can be introduced as a complement to other clustering methods. In addition, this method can optimally reduce the number of class members and classify large datasets. This method can be considered a complement to other clustering approaches. In addition, this method can reduce the size of the data set by applying previous classifications, although large data sets can also be clustered (Zahmatkesh et al. 2021). The k-means method is known for its relatively simple implementation and acceptable results. However, a direct algorithm of the k-means method requires significant time to product the number of vectors and clusters per iteration, especially for large data sets. The k-means algorithm, despite being a simple classification method, shows an acceptable performance. (Sidqi and Kakbra 2014).

K-means can be considered as an optimization issue to reduce the clustering error of a target. The purpose of the k-means algorithm is to optimize and minimize the objective function, which represents the square error function (MacQueen 1967):

$$J = \sum_{j=0}^{k} \sum_{i=0}^{n} \left| X_i^{(j)} - C_j \right|^2 \tag{7}$$

where J, k, n, X, and C is the objective function, the number of clusters, the number of data points, the data points, and the center of clusters, respectively. In this method, the data subsets are identified by a center, and the data points are assigned to clusters based on their similarities (Euclidean distance from their center of mass) which are often determined after data partitioning (McCreery and Al-Mudhafar 2017).

Machine learning methods

The use of neural networks in various branches of engineering is increasing, so that knowing how it works and how to use it is essential for petroleum engineers. In the following, while introducing the structure and operation of some of the most important machine learning methods, some of its applications in petroleum engineering are mentioned.

Several studies have presented the use of artificial intelligence in the petroleum engineering (Dougherty 1972; Braswell 2013, Kuang et al. 2021). In the oil upstream industry, the use of machine learning and optimization methods could be divided into the following three categories:

a) Exploration

  i. Determination of petrophysical parameters (Kiran and Salehi 2020, Mohammadian et al. 2022)
  ii. Geophysical processing and interpretation (Wang et al. 2018)
  iii. Determination of geomechanical parameters (Ebrahimi et al. 2022, Syed et al. 2022)
  iv. Determination and interpretation of well survey charts and wireline logs (Bestagini et al. 2017, Akinnikawe et al. 2018)
  v. Constructing a static and geological model of the reservoir (Bai and Tahmasebi 2020, Otchere et al. 2021)
  vi. etc.

b) Development and knowledge of reservoir and field

  i. Upscaling (Menke et al. 2021, Wang et al. 2022)
  ii. Preparation of (Sircar et al. 2021), and well operation (Junior et al. 2022)
  iii. Improving drilling operations (Bello et al. 2015; Noshi and Schubert 2018)
  iv. Improving reservoir simulation (Wang et al. 2020, Samnioti et al. 2022)
  v. History matching (Jo et al. 2021, Srinivasan et al. 2021)
  vi. Characterizing reservoir fluid (Ramirez et al. 2017; Onwuchekwa 2018)
  vii. Enhanced oil recovery screening (Cheraghi et al. 2021, Pirizadeh et al. 2021)
  viii. etc.

c) Field production

  i. Improvement of extraction operations (Teixeira and Secchi 2019, Pandey et al. 2021)
  ii. Maintenance of down hole pumps (Bangert and Sharaf 2019; Bangert 2021)
  iii. Improvement of artificial lift system (Syed et al. 2022)
  iv. Improvement of injection operations (Artun 2020, He et al. 2021)
  v. Improvement of well stimulation operations (Wang and Chen 2019, Liu et al. 2022)
  vi. Improvement of hydraulic fracturing operation (Morozov et al. 2020)
  vii. Arranging the pipelines (Soomro et al. 2022)
  viii. Smart well completion operation and well pattern (Castiñeira et al. 2018)
  ix. etc.

Data-driven methods show engineers a path that enables them to quickly confirm well and field performance in a very short period of time. Machine learning models such as artificial neural networks are definitely not a substitute for conventional methods such as numerical simulation; instead, a hybrid approach of machine learning modeling can provide more reliable results.

SVM method

The SVM method, which originates from statistical theory, is one of the supervised learning methods. This method is used for classification and regression (Noble 2006). The goal of this method is to find the hyper-plane that has the greatest distance from the data of the two classes (Reynolds 2001). This goal is achieved by training the SVM algorithm by means of a set of data (VAPNIK et al. 1998).

Support vectors are essentially a set of points in the n-dimensional space on which the boundary of the classes is determined. That is, by moving one of them, the output of the classification may change (Üstün et al. 2005). If each data is represented by x_i and has the number of D attributes which are labeled with specific values of y_i, Eq. 8 can be written as follows (Boser et al. 1992):

$$D = \left((x_i, y_i) | x \in R^d, y_i \in R, i = 1, 2, \ldots, n \right) \tag{8}$$

The purpose of this algorithm is to find the equation between input and output data, which is defined in Eq. 9:

$$f(x) = W^T . \emptyset(X) + b \tag{9}$$

where W and b are weight vectors and bias values, respectively. The SVM is a linear regression whose dimensions are the number of data attitudes. This algorithm tries to reduce the complexity of the model by minimizing $\|W\|^2$. In this algorithm, the objective function is defined by Eqs. 10 and 11:

$$\min \frac{1}{2} W^T W + C \sum_{i=1}^{N} (\varepsilon_i + \varepsilon_i^*) \tag{10}$$

$$\begin{cases} y_i - \left(W^T . \emptyset(X) + b \right) \le \varepsilon + \varepsilon_i \\ \left(W^T . \emptyset(X) + b \right) - y_i \le \varepsilon + \varepsilon_i^* \\ \varepsilon_i, \varepsilon_i^* \ge 0, i = 1, 2, \ldots N \end{cases} \tag{11}$$

where ε_i and ε_i^* are ineffective variables for target values which are less and greater than ε, respectively. C is used to balance model complexity and training error (Alonso et al. 2013, Mehdizadeh et al. 2014). Figure 3 shows the SVM hyper-plane for a sample data.

LogitBoost method

Boosting algorithms was originally proposed to combine several weak classifiers together to improve the classification performance. LogitBoost is additive logistic regression model. This algorithm, which is a subset of the meta-learning algorithm, is a modified model of the AdaBoost algorithm. This algorithm is introduced by Friedman et al. which uses incorrect classifications of previous models and creates a new classification class with higher accuracy (Friedman et al. 2000, Peng and Chiang 2011, Fakhraei et al. 2014). The AdaBoost algorithm uses a binomial probability logarithm to change the number function linearly. For this reason, it has limitations in noise management. The LogitBoost model is like the Ada Boost model. The main idea behind LogitBoost is to apply boosting in building a Logitmodel. The LogitBoost algorithm is designed to solve this problem (Friedman et al. 2000). The AdaBoost algorithm is more popular for classification, but the LogitBoost algorithm performs better in outbound data. Readers are referred to study of Friedman et al. for further reading on classification steps in LogitBoost algorithm.

LogitBoost is designated as a "weak" or "basic" learning algorithm. LogitBoost iteratively takes different training examples because the base learning algorithm generates a new weak prediction rule, which causes many rounds, and the subsequent boosting algorithm must transform these weak rules into a strong prediction rule, which is usually much more accurate than weak prediction (Friedman et al. 2000).

ANN method

The structure of ANN is similar to the biological network of the human body. This network can extend to imitate the function of the human brain in some way (Shepherd 1990). This algorithm is made up of artificial neurons which are the smallest unit of data processing (Sengel and Turkarslan 2020).

ANN are usually composed of several layers, which are known as input, hidden, and output layers. This network follows complex mathematical equations. These mathematical equations make connections between neurons and weights. It also optimizes network weights to achieve an optimal output. Each of the neurons processes the inputs to produce the outputs (Rezrazi et al. 2016).

In this algorithm, a random weight is determined for calculating the output for each neuron:

$$\text{Output}_i = W_{ij} X_i \tag{12}$$

where X_i and W_{ij} are the input and the weight, respectively. The output of neurons is the corresponding neuron input
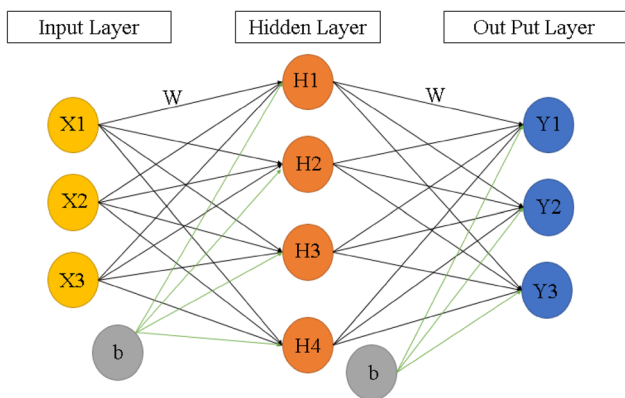
Fig. 4 Artificial neural network

in the next layer. After generating the first output layer, the activation function is applied to all of them. There are different types of activation functions, but the most common in classification is the sigmoid activation function, which is calculated by Eq. 13 (Okon et al. 2021):

$$F\left(Output_i\right) = \frac{1}{1 + e^{-Output_i}} \tag{13}$$

where $F(Output_i)$ is the value of the sigmoid function and $Output_i$ is the output layers. The result of this activation function is 0 or 1, which indicates whether each neuron is active or inactive. By this way, the classification of data is done. Figure 4 represents a multilayer neural network.

RF method

The RF algorithm was developed by Breiman in 2001. The RF algorithm has been extensively used in prediction and classification. It is a hybrid machine learning algorithm and tree-based classifier (Breiman 2001, Liu et al. 2012, Biau and Scornet 2016). This algorithm consists of a combination of tree predictors. Each tree makes a single choice for the most desirable classification in combination with a set of classified trees, and then the final result is given by combining these results. RF fits many classification trees to a data set, and then combines the predictions from all the trees. This algorithm, due to its high precision in classification, detects remote well data and separates it from the original data. RF algorithm is consist of a set of structured tree classifiers h(x,k), which is not dependent on random distribution vectors and each tree gives a single choice for the most desirable classification at the x input (Kumar et al. 2016). The kth tree is shown as θ_k, and each tree is set and distributed evenly and independently based on a set of training samples and a random variable in the Breiman RF model. Therefore,

to create a classification of more than one system of classifier h (x, θ_k) in which x is the input vector after k load, the classifier sequences $h_1(x)$, $h_2(x)$…,$H_k(x)$ is obtained. The final result of this system is chosen by a majority vote. The decision function is shown in Eq. 14 (Liu et al. 2012):

$$H(x) = \arg\max \sum_{i=1}^{k} I\left(h_i(x) = Y\right) \tag{14}$$

where (x),h_i, Y and $I(h_i(x) = Y)$ is a combination of the classification model, decision tree model, output variable, and pointer function, respectively. In the RF algorithm, selecting the best classification result for a given input variable is such that each tree has the right to vote for the most desirable outcome. Figure 5 shows the schematic of RF method structure.

Logistic regression method

A statistician named Galton used regression for the first time to describe his observations in the nineteenth century. Carl Pearson developed regression as a mathematical basis and used it to express the relationship between two quantities (Anderson et al. 2003). Logistic regression expresses the odds ratio of a variable in the presence of several explanatory variables. Multivariate logistic regression is a statistical technique that is used to estimate the probability of the output of variables. For example, the presence or absence of death (Sperandei 2014). Independent variables affect the probability of occurrence of the dependent variable (Anderson et al. 2003). The logarithm of chance is modeled as shown in Eq. 15 (Sperandei 2014):

$$Log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m \tag{15}$$

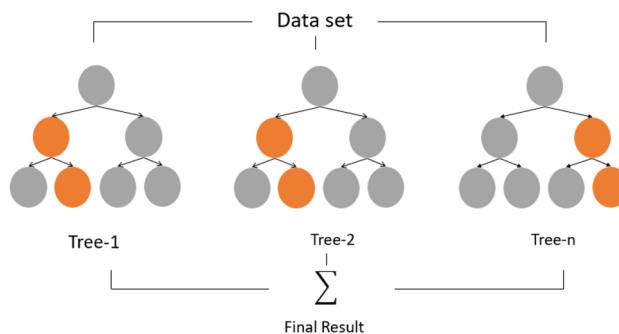**Fig. 5** Schematic of RF method structure

where π is the probability of an event, β_i are the regression coefficients associated with the reference group and the explanatory variables x_i. The reference group is denoted by β_0, and β_0 is formed by the members that represent the reference level of each of the variables($x_{1...m}$). In addition to the above explanations, the logistic regression equation is presented in another form, which is shown in Eq. 16 according to the article (Cramer 2002).

$$P(z) = \frac{\exp(Z)}{1 + \exp(Z)} \qquad (16)$$

P is similar to the density distribution function symmetric to the midpoint of zero, Z is an integer, and the P value is between 0 and 1.

Application of machine learning and optimization methods in petroleum engineering

Comparison of machine learning algorithms

As mentioned, the algorithms used in this study include SVM, LogitBoost, ANN, RF, and logistic regression. These algorithms are among the most common algorithms used in petroleum engineering. Table 3 summarizes the advantages and disadvantages of each method.

Fig. 6 RQI diagram in φ_Z for all sampled areas

Results and discussion

Petrophysical properties of the reservoir have been collected using 212 core data and several petrophysical well logs in an oil field southwest Iran. In this oil field, the reservoir porosity and permeability are varied from 2.1% to 34.1% and 0.1 to 44.3 mm at depth of 2157 to 2393.2 m, respectively. The field flow units are estimated using conventional methods including Winland R35, FZI, DRT, and k-means. These methods determine the number of flow units using core data (porosity and permeability).

By implementing the FZI method on the used data, the studied field has four flow units. Figure 6 shows FZI points that are on a line and have similar pore characteristics.

Table 3 Summary of the advantages and disadvantages of each of the methods used in this article (Haghighat et al. 2013; Mohamed 2017; Kour and Gondhi 2020)

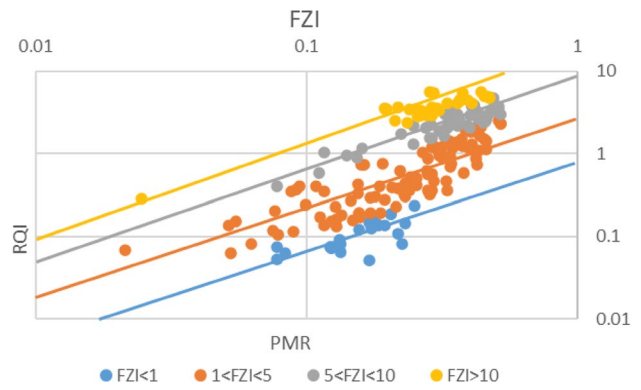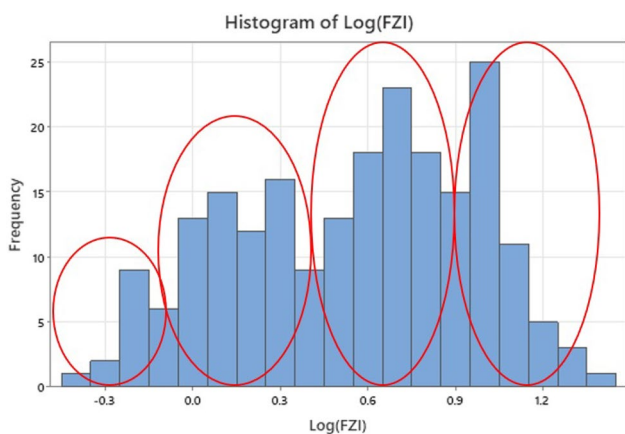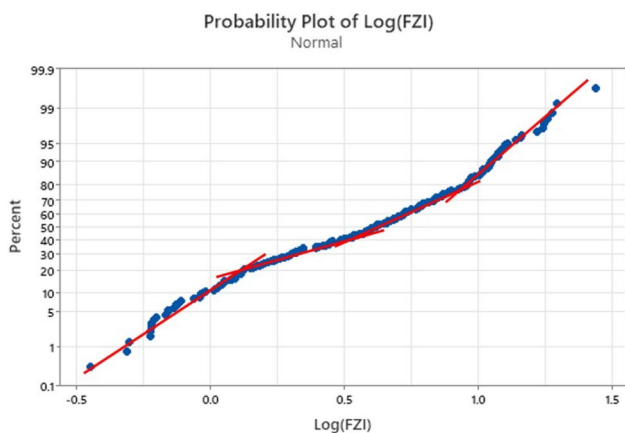| Method	Advantages	Disadvantages
SVM	- High accuracy - Working well in high-dimensional space - Using a subset of learning points, thus using much less memory	- Having a long learning time - Not suitable in practice for large data sets - Do not work well if there is overlap in classes
LogitBoost	- Boosting is basically a group model, so its predictions are easy to interpret - Strong predictive power - Resistant to the overfitting	- Sensitive to noises - Difficult in scaling
ANN	- High efficiency in performing activities - Fast learning tasks - Storing information on the entire network - Ability to provide the data to be processed in parallel	- Hardware dependence - Difficulty of showing the problem to the network - Requiring high processing time for big neural networks
RF	- Reducing the overfitting in decision trees and help to improve accuracy - flexible to both classification and regression problems - Automating the missing values in the data - Normalizing of data is not required	- Requiring the much computational power and resources - Building numerous trees to combine their outputs - Requiring much time for training - Suffering from interpretability and failing in determine the significance of each variable
Logistic regression	- Easy to implement, interpret, and very efficient to train - Very fast at classifying unknown records - Less inclined to overfitting - Very fast at classifying unknown records	- Nonlinear problems cannot be solved with logistic regression - Can overfit in high-dimensional datasets - Using only to predict discrete functions - Assume the linearity between the dependent variable and the independent variables

Fig. 7 FZI logarithm histogram

Fig. 8 Cumulative probability diagrams to determine FZI boundaries

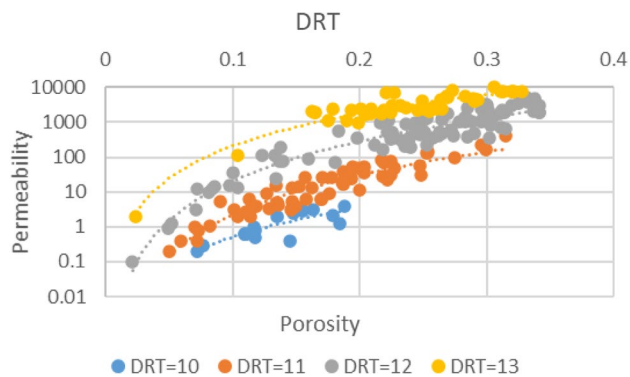**Fig. 9** Kazhdumi HFU classification results based on Winland R35 method

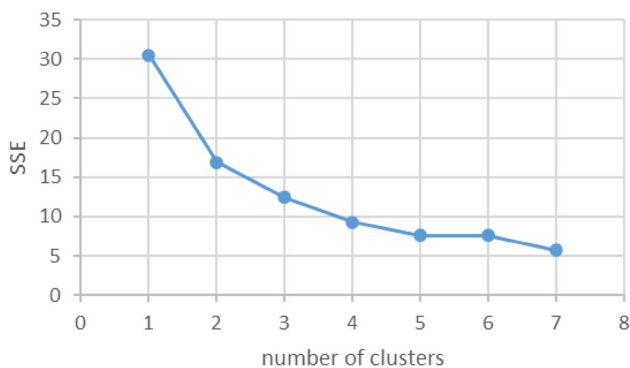**Fig. 10** Kazhdumi HFU classification results based on DRT method

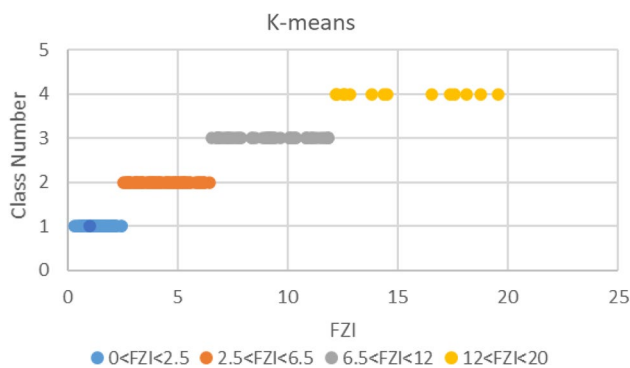**Fig. 11** Reduction in sum square error by increasing the number of clusters

Fig. 12 Kazhdumi HFU classification results based on k-means algorithm

Figure 7 shows four set of data represented by a multidimensional distribution for case study data. By plotting the cumulative probability diagram, the number of flow units could be determined by data trend breaks. According to Fig. 8, four flow units were obtained by this diagram.

Based on this study, the Winland R35 method indicates the presence of four HFUs in the Kazhdumi reservoir (Fig. 9).

The results of the DRT method on the studied field data are shown in Fig. 10 which suggests four HFUs for the Kazhdumi reservoir.

Results of applying k-means on available data show that the minimum square error decreases with increasing number of categories. However, increasing the number of categories by more than four has no significant effect on reducing the minimum square error. Figure 11 shows the process of reducing the least squares of error. As a result, four HFUs were considered for the studied reservoir by this method (Fig. 12).

Access to core data (porosity and permeability) is only possible through the core and in the laboratory. Also, the core data is not available for all wells and at all depths. To overcome this issue, in this study, petrophysical logs data which are available in most wells including RT, DT, HCAL, NPHI, RHOZ, and PEFZ have been used as input parameters to machine learning methods. Using artificial intelligence algorithms, the number of flow units could be calculated.

For this purpose, the number of flow units calculated by the FZI method is considered as target values. Log data corresponding to the data depth used in the FZI method is preprocessed as input data. SVM, LogitBoost, ANN, RF, and logistic regression algorithms are trained by 70% of the input data, and 30% of the data is used for testing. The number of field flow units is classified by petrophysical log data. The accuracy of the algorithms used is measured according to Eq. 17:

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{17}$$

To select the best number of training data, the performance of different algorithms was measured in different percentages of train/test data, and finally, 70% was selected as the optimal number of training data. The performance of the algorithms is shown in Fig. 13.

Based on available data in this study, at the first, the data is normalized, then divided into two categories of train and test with ratio of 70/30. The SVM algorithm with linear kernel function classified all data (including train and test data) with 90.46% accuracy. The confusion matrix of this algorithm for test data is shown in Fig. 14 with accuracy of 85.94%.
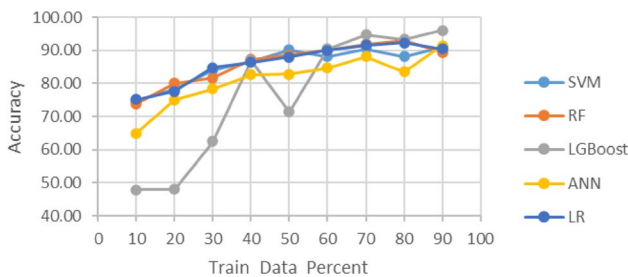

	Class 1	Class 2	Class 3	Class 4
Class 1	3	4	0	0
Class 2	1	29	1	0
Class 3	0	1	15	0
Class 4	0	1	1	8

Fig. 14 Confusion matrix of support vector machine algorithm applied to data

Applying the LogitBoost method to available data in this study for 70/30 ratio of test to train data shows 94.84% and 95.31% accuracy for the classification of all data and test data, respectively. Figure 15 illustrates the confusion matrix of classification.

The result of ANN algorithm in this study for 70/30 ratio of test to train data shows 88.12% and 73.44% accuracy for the classification of all data and test data, respectively. Figure 16 demonstrates the ANN confusion matrix classification.

In this study, RF algorithm shows 91.87% accuracy in classification using 70% and 30% of all data as train and test, respectively. It also shows 90.63% accuracy for the

Fig. 13 Accuracy of the algorithms with different percentages of training data

	Class 1	Class 2	Class 3	Class 4
Class 1	7	0	0	0
Class 2	1	30	0	0
Class 3	0	0	16	0
Class 4	1	1	0	8

Fig. 15 Confusion matrix of LogitBoost algorithm applied to data

	Class 1	Class 2	Class 3	Class 4
Class 1	4	3	0	0
Class 2	7	24	0	0
Class 3	0	3	13	0
Class 4	1	1	2	6

Actual classes (vertical label), Predicted classes (horizontal label)

Fig. 16 Confusion matrix of ANN algorithm applied to data

	Class 1	Class 2	Class 3	Class 4
Class 1	7	0	0	0
Class 2	1	30	0	0
Class 3	0	0	16	0
Class 4	1	1	0	8

Actual classes (vertical label), Predicted classes (horizontal label)

Fig. 18 Confusion matrix of logistic regression algorithm applied to data

	Class 1	Class 2	Class 3	Class 4
Class 1	7	0	0	0
Class 2	0	30	0	1
Class 3	0	1	12	3
Class 4	0	1	0	9

Actual classes (vertical label), Predicted classes (horizontal label)

Fig. 17 Confusion matrix of random forest algorithm applied to data

Fig. 19 Accuracy of different machine learning algorithm used in this study

better than other algorithms in classification with an accuracy of 94.84%.

classification of test data. The confusion matrix of this algorithm is shown in Fig. 17.

Logistic regression algorithm has reached 91.56% accuracy with 30% test data and 70% training data in this study. Figure 18 shows the confusion matrix of logistic regression algorithm. Based on this confusion matrix, logistic regression algorithm shows 95.31% accuracy for the classification of test data.

Finally, machine learning methods are compared to select the best method used. The results obtained from the best performance of each algorithm are shown in Fig. 19. As shown in this figure, the LogitBoost method performed
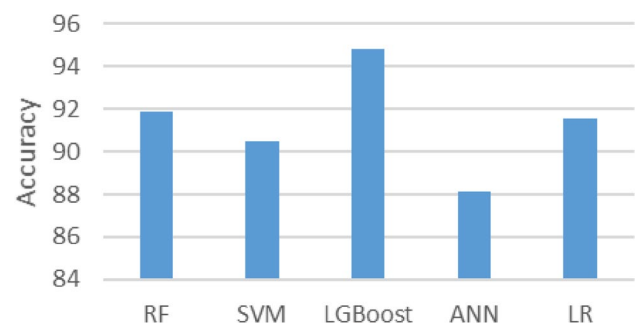
Conclusions

In this article, a variety of conventional methods and machine learning algorithms were investigated in determining hydraulic flow units (HFUs), and the performance of each method was evaluated. The following results, which also flesh out the innovative nature of the work, are as follows:

– The k-means method and the use of sum of squares error (SSE) operate independently of the user and show an effective performance for determining the optimal number of HFUs, and it is also fully consistent with the Flow Zone Index (FZI) method.

– Conventional methods determine flow units only by using porosity and permeability parameters obtained from core analysis, while these data may not be available along the entire length of the reservoir. Therefore, the use of intelligent data-driven methods that use petrophysical log data can be a suitable alternative to conventional methods, especially in wells where core data is not available.

– In this article, support vector machine (SVM), artificial neural network (ANN), random forest (RF), LogitBoost (LB), and logistic regression (LR) machine learning methods were used to determine flow units using petrophysical logs. The results showed that among the different machine learning algorithms used in this study, the LB method has the best performance in determining HFUs. After that, RF, LR, SVM and ANN methods have the best accuracies, respectively.

Declarations

References

Abed AA (2014) Hydraulic flow units and permeability prediction in a carbonate reservoir, Southern Iraq from well log data using non-parametric correlation. Int J Enhan Res Sci Technol 3(1):480–486

Abnavi ADT, Amir Karimian; Qajar, Jafar, (2021) Hydraulic flow units and ANFIS methods to predict permeability in heterogeneous carbonate reservoir: Middle East gas reservoir. Arab J Geosci 14(9):1–10

Abnavi, A. D., et al. (2018). Identification of Hydraulic Flow Units in a heterogeneous carbonate reservoir: case study one of the Iranian gas fields. 2nd International Biennial Oil, Gas and Petrochemical Conference.

Akinnikawe, O., et al. (2018). Synthetic well log generation using machine learning techniques. SPE/AAPG/SEG unconventional resources technology conference, OnePetro. https://doi.org/10.15530/URTEC-2018-2877021

Al-Ajmi FA and SA Holditch (2000) Permeability estimation using hydraulic flow units in a central Arabia reservoir. SPE annual technical conference and exhibition, OnePetro. https://doi.org/10.2118/63254-MS

Alonso J et al (2013) Support vector regression to predict carcass weight in beef cattle in advance of the slaughter. Comput Electron Agric 91:116–120

Amaefule JO et al. (1993) Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/wells. SPE annual technical conference and exhibition, OnePetro. https://doi.org/10.2118/26436-MS

Anderson RP et al (2003) Understanding logistic regression analysis in clinical reports: an introduction. Ann Thorac Surg 75(3):753–757

Artun E (2020) Performance assessment and forecasting of cyclic gas injection into a hydraulically fractured well using data analytics and machine learning. J Petrol Sci Eng 195:107768

Bai T, Tahmasebi P (2020) Hybrid geological modeling: combining machine learning and multiple-point statistics. Comput Geosci 142:104519

Bangert P, Sharaf S (2019) Predictive maintenance for rod pumps. SPE Western Regional Meeting, OnePetro. https://doi.org/10.2118/195295-MS

Bangert P (2021) Predictive and diagnostic maintenance for rod pumps. machine learning and data science in the oil and gas industry, Elsevier,pp 237–252.

Bello O et al (2015) Application of artificial intelligence methods in drilling system design and operations a review of the state of the art. J Artif Intell Soft Comput Res 5(2):121–139

Bestagini P et al. (2017) A machine learning approach to facies classification using well logs. Seg technical program expanded abstracts 2017 Society of Exploration Geophysicists pp 2137–2142.

Biau G, Scornet E (2016) A random forest guided tour. TEST 25(2):197–227

Boser BE et al. (1992) A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory.

Braswell G (2013) Artificial Intelligence Comes of Age in Oil and Gas. J Petrol Technol 65(01):50–57

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Castiñeira, D., et al. (2018). Machine learning and natural language processing for automated analysis of drilling and completion data. SPE Kingdom of Saudi Arabia annual technical symposium and exhibition, OnePetro https://doi.org/10.2118/192280-MS

Chekani M, Kharrat R (2012) An integrated reservoir characterization analysis in a carbonate reservoir: a case study. Pet Sci Technol 30(14):1468–1485

Cheraghi Y et al (2021) Application of machine learning techniques for selecting the most suitable enhanced oil recovery method; challenges and opportunities. J Petrol Sci Eng 205:108761

Cramer JS (2002) The origins of logistic regression.

Dougherty E (1972). Application of optimization methods to oilfield problems: proved, probable, possible. fall meeting of the society of petroleum engineers of AIME, OnePetro. https://doi.org/10.2118/3978-MS

Ebanks Jr W (1987) Flow unit concept-integrated approach to reservoir description for engineering projects. AAPG (Am. Assoc. Pet. Geol.) Bull. (United States) 71(CONF-870606-).

Ebrahimi A et al (2022) Estimation of shear wave velocity in an Iranian oil reservoir using machine learning methods. J Petrol Sci Eng 209:109841

Elnaggar OM (2018) A new processing for improving permeability prediction of hydraulic flow units, Nubian Sandstone, Eastern Desert, Egypt. J Petroleum Explor Prod Technol 8(3):677–683

El-Sayed AMA et al (2021) Rock typing based on hydraulic and electric flow units for reservoir characterization of Nubia Sandstone, southwest Sinai, Egypt. J Petroleum Explor Prod Technol 11(8):3225–3237

Fakhraei S et al (2014) Bias and stability of single variable classifiers for feature ranking and selection. Expert Syst Appl 41(15):6945–6958

Friedman J et al (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann Stat 28(2):337–407

Ghadami N et al (2015) Consistent porosity–permeability modeling, reservoir rock typing and hydraulic flow unitization in a giant carbonate reservoir. J Petrol Sci Eng 131:58–69

Gomes JS et al (2008) Carbonate reservoir rock typing-the link between geology and SCAL. Abu Dhabi International Petroleum Exhibition and Conference, OnePetro. https://doi.org/10.2118/118284-MS

Gunter G et al (1997) Early determination of reservoir flow units using an integrated petrophysical method. SPE Annual Technical Conference and Exhibition, OnePetro. https://doi.org/10.2118/38679-MS

Guo G et al (2005) Rock typing as an effective tool for permeability and water-saturation modeling: a case study in a clastic reservoir in the Oriente Basin. SPE Annual Technical Conference and Exhibition, OnePetro. https://doi.org/10.2118/97033-PA

Haghighat M et al (2013) A review of data mining techniques for result prediction in sports. Adv Comput Sci 2(5):7–12

He, X., et al. (2021). Gas Injection optimization under uncertainty in subsurface reservoirs: an integrated machine learning-assisted workflow. ARMA/DGS/SEG International Geomechanics Symposium, OnePetro.

Hosseini Bidgoli SH et al (2014) Identifying flow units using an artificial neural network approach optimized by the imperialist competitive algorithm. Iranian J Oil Gas Sci Technol 3(3):11–25

Jo H et al (2021) Machine learning assisted history matching for a deepwater lobe system. J Petrol Sci Eng 207:109086

Junior JRB et al (2022) A comparison of machine learning surrogate models for net present value prediction from well placement binary data. J Petrol Sci Eng 208:109208

Khadem B et al (2020) Integration of rock physics and seismic inversion for rock typing and flow unit analysis: a case study. Geophys Prospect 68(5):1613–1632

Kiran R and S Salehi (2020) Assessing the relation between petrophysical and operational parameters in geothermal wells: a machine learning approach. In: Proceedings of the 45th Workshop on Geothermal Reservoir Engineering, Stanford, CA, USA.

Kolodzie, S. (1980). Analysis of pore throat size and use of the Waxman-Smits equation to determine OOIP in Spindle Field, Colorado. SPE annual technical conference and exhibition, OnePetro. https://doi.org/10.2118/9382-MS

Machine learning techniques: a survey. In: International conference on innovative data communication technologies and application, Springer. https://doi.org/10.1007/978-3-030-38040-3_31

Kuang L et al (2021) Application and development trend of artificial intelligence in petroleum exploration and development. Pet Explor Dev 48(1):1–14

Kumar R et al (2016) A methodology of porosity estimation from inversion of post-stack seismic data. J Nat Gas Sci Eng 28:356–364

Liu, H., et al. (2022). Using machine learning method to optimize well stimulation design in heterogeneous naturally fractured tight reservoirs. SPE Canadian energy technology conference, OnePetro. https://doi.org/10.2118/208971-MS

Liu Y et al. (2012) New machine learning algorithm: random forest. Information Computing and Applications, Berlin, Heidelberg, Springer Berlin Heidelberg.

MacQueen J (1967) Some methods for classification and analysis of multivariate observations.In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA.

McCreery E and W Al-Mudhafar (2017) Geostatistical Classification of Lithology Using Partitioning Algorithms on Well Log Data-A Case Study in Forest Hill Oil Field, East Texas Basin. In: 79th EAGE Conference and Exhibition 2017, European Association of Geoscientists & Engineers.

Mehdizadeh SA et al (2014) Early determination of Pharaoh Quail sex after hatching using machine vision. Bull Env Pharmacol Life Sci 3:05–11

Menke HP et al (2021) Upscaling the porosity–permeability relationship of a microporous carbonate for Darcy-scale flow with machine learning. Sci Rep 11(1):2625

Mohamed AE (2017) Comparative study of four supervised machine learning techniques for classification. Int J Appl 7(2):1–15

Mohammadian E et al (2022) A case study of petrophysical rock typing and permeability prediction using machine learning in a heterogenous carbonate reservoir in Iran. Sci Rep 12(1):4505

Mohebian R et al (2013) Channel detection using instantaneous spectral attributes in one of the SW Iran oil fields. Bollettino Di Geofisica Teorica Ed Applicata 54:271–282

Morozov A et al. (2020) Machine Learning on Field Data for Hydraulic Fracturing Design Optimization: Digital Database and Production Forecast Model. First EAGE Digitalization Conference and Exhibition, EAGE Publications BV.

Motiei H (1995) Petroleum geology of zagros. Geological Survey of Iran, Farsi.

Nabawy BS et al (2018) Petrophysical and microfacies analysis as a tool for reservoir rock typing and modeling: Rudeis Formation, off-shore October Oil Field, Sinai. Mar Pet Geol 97:260–276

Nairn A, Alsharhan A (1997) Sedimentary basins and petroleum geology of the Middle East. Elsevier

Nayak NP et al (2021) Application of hydraulic flow unit for prediction of flow zone in carbonate reservoir. Arab J Geosci 14(8):1–8

Noble WS (2006) What is a support vector machine? Nat Biotechnol 24(12):1565–1567

Noshi CI, Schubert JJ (2018) The role of machine learning in drilling operations; a review. SPE/AAPG Eastern Regional Meeting, OnePetro. https://doi.org/10.2118/191823-18ERM-MS

Nouri-Taleghani M et al (2015) Determining hydraulic flow units using a hybrid neural network and multi-resolution graph-based clustering method: case study from south pars gasfield, Iran. J Pet Geol 38(2):177–191

Okon AN et al (2021) Artificial neural network model for reservoir petrophysical properties: porosity permeability and water saturation prediction. Modeling Earth Syst Environ 7(4):2373–2390

Onwuchekwa C (2018) Application of machine learning ideas to reservoir fluid properties estimation. SPE Nigeria Annual International Conference and Exhibition, OnePetro. https://doi.org/10.2118/193461-MS

Otchere DA et al (2021) Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. J Petrol Sci Eng 200:108182

Pandey RK et al (2021) Identifying applications of machine learning and data analytics based approaches for optimization of upstream petroleum operations. Energ Technol 9(1):2000749

Peng, H.-W. and P.-J. Chiang (2011). Control of mechatronics systems: Ball bearing fault diagnosis using machine learning techniques. 2011 8th Asian Control Conference (ASCC), IEEE.

Pirizadeh M et al (2021) A new machine learning ensemble model for class imbalance problem of screening enhanced oil recovery methods. J Petrol Sci Eng 198:108214

Rabbani A (2007) Petroleum geochemistry, offshore SE Iran. Geochem Int 45(11):1164–1172

Rabbani A (2013) Geology and geochemistry of Persian Gulf oil. Tafresh University Publications, Tafresh

Ramirez, A., et al. (2017). Prediction of PVT properties in crude oil using machine learning techniques MLT. SPE Latin America and caribbean petroleum engineering conference, OnePetro. https://doi.org/10.2118/185536-MS

Reynolds DA (2001) Automatic speaker recognition: Current approaches and future trends. Speaker Verification: from Research to Reality 5:14–15

Rezrazi A et al (2016) An optimisation methodology of artificial neural network models for predicting solar radiation: a case study. J Theor Appl Climatol 123(3–4):769–783

Riazi Z (2018) Application of integrated rock typing and flow units identification methods for an Iranian carbonate reservoir. J Petrol Sci Eng 160:483–497

Samnioti A et al (2022) Application of machine learning to accelerate gas condensate reservoir simulation. Clean Technol 4(1):153–173

Sengel A, Turkarslan G (2020) Assisted history matching of a highly heterogeneous carbonate reservoir using hydraulic flow units and artificial neural networks. SPE Europec, OnePetro. https://doi.org/10.2118/200541-MS

Shalaby MR (2021) Petrophysical characteristics and hydraulic flow units of reservoir rocks: Case study from the Khatatba Formation, Qasr field, North Western Desert, Egypt. J Petrol Sci Eng 198:108143

Sharifi-Yazdi M et al (2020) Diagenetic impacts on hydraulic flow unit properties: insight from the Jurassic carbonate Upper Arab Formation in the Persian Gulf. J Petroleum Explor Prod Technol 10(5):1783–1802

Shepherd GM (1990) Introduction to synaptic circuits. The synaptic organization of the brain: 3–31.

Sidqi HM, Kakbra JF (2014) Image classification using K–mean algorithm. Int J Emerg Trends Technol Comput Sci 3:38–41

Siebold E (1969) Bodengestaldes des Persischen Golfs'.Meteor' Forsch. Ergebn. Gebruder Borntrageer, Berlin 2:29–56

Sircar A et al (2021) Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Res 6(4):379–391

Soleymanzadeh A et al (2019) Effect of overburden pressure on determination of reservoir rock types using RQI/FZI, FZI* and Winland methods in carbonate rocks. Petroleum Sci 16(6):1403–1416

Soomro AA et al (2022) Integrity assessment of corroded oil and gas pipelines using machine learning: A systematic review. Eng Fail Anal 131:105810

Spearing M et al. (2001) Review of the Winland R35 method for net pay definition and its application in low permeability sands. In: Proceedings of the 2001 International Symposium of the Society of Core Analysts, Citeseer.

Sperandei S (2014) Understanding logistic regression analysis. Biochemia Medica 24(1):12–18

Srinivasan S et al (2021) A machine learning framework for rapid forecasting and history matching in unconventional reservoirs. Sci Rep 11(1):21730

Syed FI et al (2022) Application of ML & AI to model petrophysical and geomechanical properties of shale reservoirs–A systematic literature review. Petroleum 8(2):158–166

Teixeira AF, Secchi AR (2019) Machine learning models to support reservoir production optimization. IFAC-PapersOnLine 52(1):498–501

Tiab, D. and E. C. Donaldson (2015). Petrophysics: theory and practice of measuring reservoir rock and fluid transport properties, Gulf professional publishing.

Üstün B et al (2005) Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. Anal Chim Acta 544(1–2):292–305

Vapnik VN et al (1998) Statistical Learning Theory. Wiley, New York

Wang S, Chen S (2019) Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. J Petrol Sci Eng 174:682–695

Wang Z et al (2018) Successful leveraging of image processing and machine learning in seismic structural interpretation: a review. Lead Edge 37(6):451–461

Wang K et al (2020) Practical application of machine learning on fast phase equilibrium calculations in compositional reservoir simulations. J Comput Phys 401:109013

Wang Y et al (2022) Machine learning assisted relative permeability upscaling for uncertainty quantification. Energy 245:123284

Winland H. J. A. P. R. R. N. F.-G. (1972). "Oil accumulation in response to pore size changes, Weyburn field, Saskatchewan."

Wu Z et al (2016) Pore-scale experiment on blocking characteristics and EOR mechanisms of nitrogen foam for heavy oil: a 2D visualized study. Energy Fuels 30(11):9106–9113

Wu Z et al (2018) Emulsification and improved oil recovery with viscosity reducer during steam injection process for heavy oil. J Ind Eng Chem 61:348–355

Yi D et al. (2021). Research on Identification of Flow Units Based on FZI. E3S Web of Conferences, EDP Sciences.

Zahmatkesh I et al (2021) Integration of well log-derived facies and 3D seismic attributes for seismic facies mapping: A case study from mansuri oil field, SW Iran. J Pet Sci Eng 202:108563

Zargar G et al (2020) Reservoir rock properties estimation based on conventional and NMR log data using ANN-Cuckoo: A case study in one of super fields in Iran southwest. Petroleum 6(3):304–310

Zargari M et al (2013) Permeability prediction based On hydraulic flow units (HFUs) and adaptive neuro-fuzzy inference systems (ANFIS) in an Iranian Southern Oilfield. Pet Sci Technol 31(5):540–549