Fuzzy support vector regression for permeability estimation of petroleum reservoir using well logs

Nastaran Moosavi[1] · Majid Bagheri[2] · Majid Nabi-Bidhendi[2] · Reza Heidari[1]

Abstract

In this paper we propose a three-step approach to predict permeability. First, by using Electrofacies Analysis (EA), data are classified into several clusters. We take advantage of EA to overcome abrupt changes of permeability which its unpredictability prevents a machine to be learned. EA is also helpful for wells that suffer from core data. Second, fuzzy membership functions are applied on data points in each Electrofacies Log (EL). Third, Support Vector Regression (SVR) is employed to predict permeability using fuzzy clustered data for areas with core missing data. To perform this process, we applied the proposed technique on four well sets of a gas field located in South of Iran; three wells devoted to training and the fourth remained for testing operation. Seven ELs derived using Multi Regression Graph-Based Clustering (MRGC) method. MRGC is able to estimate more appropriate number of clusters without prior knowledge compared to other three algorithms for our case-study area. Then, fuzzy membership functions applied to data. Thereafter, SVR applied to both fuzzy and not-fuzzy ELs. Consequently, the predicted permeability log for both fuzzy and not-fuzzy inputs correlated to real permeability (core data obtained from plugs in laboratory) in the test well. Finally, predicted permeability for each face merged together to make an estimated permeability for the whole test well. The results show that predicted permeability obtained from application of SVR on fuzzy data (FSVR) has a notably better correlation with core data for both clusters individually and the whole data compared to SVR.

Keywords Permeability · Core data · Heterogeneous reservoirs · Well logs · Support vector regression · Fuzzy SVR

Introduction

Petrophysical assessment of hydrocarbon bearing reservoirs has an important role in reservoir studies. Permeability, as the rock capability to transfer the fluids, is a vital parameter in exploitation of oil and gas reservoirs. This parameter is also helpful to identify drainage area and to employ suitable instruments for fluid extraction (Klinkenberg 1941). Reliable permeability is achievable through laboratory measurements via core data (Glover 2000). Different approaches are available to obtain permeability in the laboratory (Kozeny 1927;

Leverett 1941; Thomeer 1960; Gan et al. 2018; Singh et al. 2020; Shen et al. 2020, Mehdipour et al. 2018). Due to the fact that determining permeability from core data is very costly and also it is insufficient for lateral distribution studies, well logs have been considered as good substitutions. They are able to provide both lateral and vertical distributions, too. Conventional methods show that permeability could be obtained from porosity logs (Ameri et al. 1993). However, this method also did not produce good results especially in areas that permeability and porosity have a poor correlation. Consequently, usage of more advanced logs became popular for permeability evaluation (Smith et al. 2014, Fiandaca et al. 2018). This method was also not appropriate because advanced logs are not always available. Over recent years, variety of artificial intelligent algorithms have been employed for modeling hydrocarbon reservoirs such as artificial neural networks (ANN), deep learning, fuzzy logic and many other machine learning algorithms (Zhang et al. 2018, Okwu et al. 2019, Sinaga et al. 2019). ANN as a popular technique work based on empirical risk

✉ Majid Bagheri
majidbagheri@ut.ac.ir

[1] Department of Earth Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran

[2] Institutes of Geophysics, University of Tehran, Tehran, Iran

minimization (ERM), meaning that model is made based on training data and finally the error is minimized (Mohaghegh 2000, Iturraran et al. 2014). In this case, the model is prone to over-fitting and it may not produce a good generalization.

Fluctuation of well log values over the depth is a sign of lithology variation in sediment formations. In areas with complex structures, we have abrupt changes of permeability which prohibits the algorithm to easily find a pattern. As a result, EA is used to overcome this problem by classifying the data points due to their similarity. EA was first introduced by Serra and Abbot (1980) and this method was first used by researchers of Pecton-Shell Company.

EA is a cheap and fast method to classify different lithologies based on the structural characteristics recognizable in well logs. This method classifies data into homogenous clusters in which every cluster contains similar particles from size, shape and texture. Each cluster is called an Electrofacies Log (EL). Investigation of ELs and integrate them with geology data is a popular method to evaluate areas with core missing data. Different approaches are available to detect formation ELs and to predict permeability; such as MRGC, agglomerative hierarchical clustering (AHC), Kohonen self-organizing map (SOM) and dynamic clustering (DC) (Khoshbakht et al. 2012, Nourani et al. 2016). Each of these methods work due to a distinct algorithm; for example, MRGC is able to find number of clusters without any prior information. But in AHC and SOM, the user guesses number of clusters based on some prior information such as petrographic studies and sediment facies. Permeability modeling using EA with different algorithms and regression techniques has been carried out by various authors (Rafik et al. 2016; Criollo et al. 2016; Al-Mudhafar 2020; Liaghat et al. 2021).

In this paper, we present SVR algorithm that is able to produce notably better models compared to other regression methods. SVR produces outputs with good generalization and less prone to over-fitting problems (Novikoff 1962), (Vapnik and Chervonenkis 1979), (Vapnik 1982, 1995) (Gunn, 1998). SVR is based on structural risk minimization (SRM) providing models with good generalization. It creates a balance between training error and test error. Good results gained from this algorithm for predicting permeability have been published recently (Al-Anazi et al. 2012), (Silva et al. 2020), (Li et al. 2021).

It must be considered that measuring data in real world without noise and outliers are inevitable. Although processing of raw data reduces noise up to a good level, omitting the whole noise is not possible. For concluding a better result, applying fuzzy membership functions on data and devoting a membership degree to each data point reduces noise remarkably. In this method each data has a weight based on its degree of membership, therefore, each of the data points can affect the learning algorithm (Lin et al. 2002), (Hasan et al. 2020). FSVR has been used in different fields of study such as economy, computer and system modeling

and electromagnetic (Zhang et al. 2010), (Rustam et al. 2019), (Hosseinzadeh et al. 2020).

Methods

Electrofacies analysis (EA)

The name "electrofacies" was first introduced by Serra and Abbott (1980). EA divides well logs into several geological parts that each part represents a geologic feature. In other words, electrofacies are petrophysical well log responses that characterize a layer and separate it from other layers (Nashawi and Malallah 2009). Combining the same quality parts from each well builds ELs so that each EL can represent one or more lithofacies (Negi et al. 2006). Facies identification plays an important role in identifying petroleum and reservoir characterization. The main difference between electrofacies and facies classification of rocks in subsurface is that electrofacies does not need to outcrops, core data and cuttings. Moreover, no information about stratigraphy or depositional map of the environment is needed (Davis 2018). In some cases, electrofacies are used as a tool to distinguish between reservoir and non-reservoir segments and also to find structure similarities in the field. As a result, electrofacies are sometimes called virtual cores (Khan Mohammadi and Sherkati 2010). There are four identification algorithms of ELs in Geolog software (Emerson Paradigm 2019): MRGC, AHC, SOM and DC. A brief review of each of them is as follow:

1. **MRGC:** this algorithm locates clusters using a multidimensional dot-pattern recognition method based on nonparametric k-nearest neighbor (KNN) and graph data representation.
2. **AHC:** this is a statistical method for finding relatively homogeneous clusters of cases based on measured characteristics. It starts with each case in a separate cluster and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left.
3. **SOM:** this algorithm is an unsupervised neural network that its main property is for producing ordered low-dimensional representations of an input data space. Typically, such input data are complex and high-dimensional with data elements being related to each other in a nonlinear fashion. The basis of the method derives from biophysical studies of the brain and "brain maps" which form internal representations of external stimuli in a spatial manner.
4. **DC:** this method classifies data according to pre-defined number of clusters. It defines central point randomly, then calculates the distance between the data point with

a center. Thereafter, each data can be assigned to the nearest center. Each center location changes iteratively until all points find their nearest neighbors.

Multi resolution graph-based modeling (MRGC)

Providing more details about MRGC algorithm, it must be mentioned that this algorithm does not need prior information about number of clusters or data structure and also it chooses number of clusters automatically. In KNN algorithm, for classification purpose, a data point gets assigned to its kth nearest neighbor. Since outliers are located farther from ordinary data points, with increasing of K, algorithm finds a better opportunity to distinguish noise and outliers. Compared to other algorithms like probability density function (PDF) that data points get assigned to areas, MRGC takes advantage of capability of more detailed classification; for instance, classifying data in cases with small size or different densities. PDF shows the probability distribution of variables. This function is due to the probable areas that a variable may exist. Outliers have less amount of likelihood and they usually accumulate farther on sides, however, distinguishing of outliers from data relying on farther sides is not that easy. Moreover, MRGC uses continuous graphs to connect data points to each other, then by utilizing heuristic laws, extra graphs are emitted and data classifies into several clusters (Ye and Rabiller 2000), (Aghchelou et al. 2012). Utilizing MRGC has several advantages which we notice some of them here as follow:

- It has the capability to classify sections of well logs that are prone for geological facies.
- It does not need to prior knowledge of data.
- It automatically detects cluster numbers.
- It is able to find clusters for even very complex structures.

Support vector regression (SVR)

SVR is a modification of support-vector machine (SVM) for regression purposes which mainly works due to maximal margin and uses Vapnik's ε-insensitivity loss function (Vapnik 1995). Since SVR concept roots in SVM, so first we start with a short explanation regarding how SVM works. SVM utilized a hyperplane (decision surface) for two purposes; firstly, to minimize the error and secondly to maximize the margin. Margin is described as the distance between the hyperplane and the nearest data points. Figure 1 shows a simple description of how SVM works which is helpful for a better understanding of SVR. As Fig. 1 shows, for partitioning of two groups of data, there are hundreds of possible lines such as L1 and L2. SVM finds the line (L3) which has a maximum distance (orange lines) from the closest data points (Support Vectors). SVM is fundamentally according
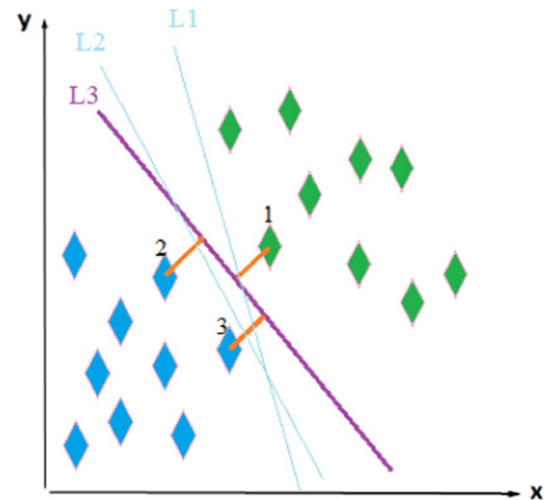

Fig. 1 Classification of two groups of data using SVM method; L1 and L2 are random possible lines which are able to classify data into two groups. L3 is optimal hyperplane that separate data due to maximizing the distance between closest data points (orange lines). Nearest data points from each group to each other are called support vectors (data points with numbers 1, 2 and 3)

to SRM. This method is the minimization of upper bound on calculated test data error. This leads to models with better generalization quality (Gunn 1998).

To calculate a linear SVR, we are supposed to have n training data points:

$$(x_1, y_1)...(x_n, y_n), \tag{1}$$

where x is the data point with response observed value y, measured for n observations.

As we know older regression methods minimized the measured error from the difference of observed values and predicted values with different approaches like least squares (minimizing sum of the squared of residuals), minimum sum method (minimizes the sum of absolute values of residuals) and Chebyshev method (minimizes the larges residual in absolute value). But SVR using ε-insensitivity loss function is a tool for measuring distance between data points and predicted function. This loss function has limitations for minimization of error with an addition second constraint. Error minimization in this method is less than or equal to an amount of ε and data points have a maximum distance of ε from the prediction function (Rustam et al. 2019). So, if we consider a linear regression form as below (Suykens et al. 2002):

$$f(x) = w^T x + b, \tag{2}$$

where w is equivalent to weight parameter and b is representative of bias.

The loss function is described as follow. It shows the space between each observed value (y) and estimated function f(x) which must be penalized:

$$(y - f(x))_\varepsilon = \begin{cases} 0 & if |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & otherwise \end{cases}. \quad (3)$$

Generally, SVR obeys two facts as follow:

$$\text{Min}_{w,b} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} \left(\xi_i + \xi_i^* \right) \quad (4)$$

subjected to:

$$\forall n : y_n - \left(Xn'w + b \right) \le \varepsilon + \xi_n \quad (5)$$

$$\forall n : \left(Xn'w + b \right) - y_n \le \varepsilon + \xi_n^*$$

$$\forall n : \xi_n^* \ge 0$$

$$\forall n : \xi_n \ge 0.$$

$M = \frac{2}{||W||}$ is the distance of support vectors from the predicted function, ξ_i and ξ_i^* are slack variables considered as all of the data points that locate outside the ε area (Fig. 2). C parameter creates a balance between slack variables and the level of model complexity.

a_i and a_i^* are positive Lagrangian multipliers which can be calculated by solving the dual optimization problem:
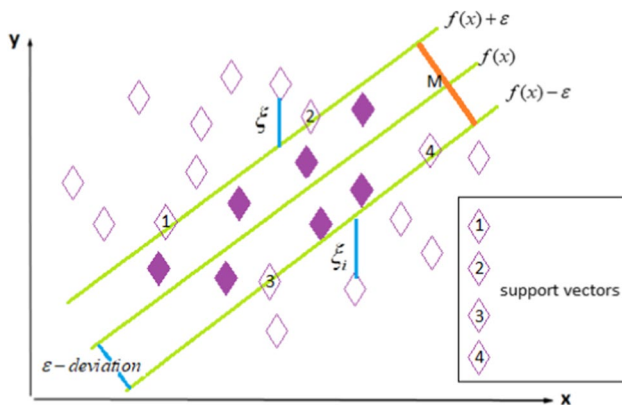

Fig. 2 Illustration of estimated function f(x) using SVR. Slack variables ξ and ξ^* indicate data points exceed ε. Their distance from green ε-tube (green tube; the area between f(x) +ε and f(x) −ε is considered as error (blue lines) that must be minimized. ε-deviation is the half distance between support vectors of two classes (M) which is supposed to be maximized

$$maximize - \frac{1}{2} \sum_{i,j=1}^{n} \left(\alpha_i - \alpha_i^* \right) \left(\alpha_j - \alpha_j^* \right) \langle X_i . X_j \rangle \quad (6)$$

$$- \varepsilon \sum_{i=1}^{n} \left(\alpha_i + \alpha_i^* \right) + \sum_{i=1}^{n} y_i \left(\alpha_i - \alpha_i^* \right)$$

$$\text{subject to} \quad \sum_{i=1}^{N} \left(\alpha_i - \alpha_i^* \right) = 0 \text{ and } \alpha_i, \alpha_i^* \in [o, c]. \quad (7)$$

By solving the mentioned problem, predicted function and *w* are obtained as follow:

$$w = \sum_{i=1}^{N} \left(a_i - a_i^* \right) X \quad (8)$$

$$f(x) = \sum_{i=1}^{n} \left(\alpha_i - \alpha_i^* \right) \langle X_i . X \rangle + b. \quad (9)$$

Fuzzy support vector regression (FSVR)

Noise and outliers are inevitable part of measured data. FSVR is one of the algorithms that has a control on data distribution and is fruitfully helpful to suppress noise in data (Bishop 2006). FSVR gives fuzzy membership values to each data point. Therefore, each data point can influence on training process differently (Rustam et al. 2019). Consequently, some data points gain more weight and some data point lose their degree of importance in regression model. Consider S_i to be a fuzzy membership value devoted to each data point in the prediction model, then $(1 - S_i)$ is a less important contribution. If the fuzzy membership value is small, then the contribution of that data point on the model is less significant (Rustam et al. 2019). Outliers and noise receive small value of membership function in FSVR so their role in predicting a function decrease. Fuzzy membership concept was first introduced by Zadeh (1965) in his first paper of fuzzy sets. Membership functions transfer data into a [0,1] interval. This means that data are mapped on a curve with different degrees and values between 0 and 1 interval by membership functions. Fuzzy is widely applicable in different fields such as face recognition (Lim et al. 2002), stock market (Rustam et al. 2019), linguistics (Le et al. 2009) and so on. In this paper we use fuzzy membership function introduced by Iin et al. (2002):

$$S_i = \begin{cases} 1 - |X_+ - X_i| / (r_+ - \delta) & \text{where } y_i = 1 \\ 1 - |X_- - X_i| / (r_- + \delta) & \text{where } y_i = -1 \end{cases}, \quad (10)$$

where S_i is fuzzy membership value devoted to each data point. X_+ is mean of data points with symbol $+1$ and X_- is the mean of data points with symbol -1.

r_+ And r_- are as follow:

$$r_+ = \max |X_+ - X_i| \quad \text{where} \quad y_i = 1 \qquad (11)$$

$$r_- = \max |X_- - X_i| \quad \text{where} \quad y_i = -1.$$

Finally, with assigning S_i value to each data point, we modify SVR to FSVR.

Figure 3 presents how fuzzy numbers can be located in a regression model. If fuzzy data are assumed to be presented as symmetric triangular numbers, then a fuzzy data $\overline{Y}_i$ can be shown by $\tilde{Y}_i = (Y_i, c)$ which Y_i is the fuzzy center and c is the fuzzy half width. Then the membership function of $\overline{Y}_i$ is expressed as follow:

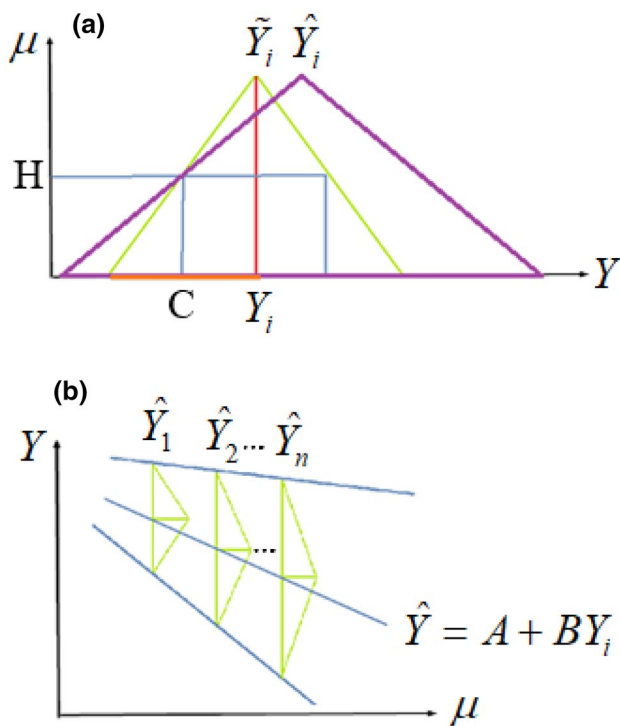$$\mu_{\overline{Y}_i}(Y) = 1 - \frac{|Y - Y_i|}{c}. \qquad (12)$$

Fig. 3 Illustration for fuzzy linear regression. **a** $\tilde{Y}_i = (Y_i, C)$ is a fuzzy data with a real number Y_i in center and half width of C. Green symmetrical triangular shows the membership function of $\tilde{Y}_i$. H is a parameter that shows how model and data are fitted properly. The purple symmetrical triangular shows the estimated fuzzy data ($\hat{Y}_i$). **b** Illustration of fuzzy linear regression model $\hat{Y} = A + BY_i$. Each green triangular is a membership function of Y_i related to its estimated fuzzy value $\hat{Y}_1$

For a linear regression $\hat{Y} = A + BY_i$ which A and B are intercept and slope coefficients, respectively. Both of these coefficients are fuzzy numbers. As a result, the estimated value of $\hat{Y}$ is also a fuzzy number. Y_i is an independent crisp number which means it does not have any uncertainty. To find a fuzzy regression model, the equation $\hat{Y} = A + BY_i$ must be solved and coefficients must be found.

H is a value that shows how good the fitness of the model is, it means how well data are fitted to the regression model. H is a number between 0 and 1. When H gets closer to 0 means that model fits the data appropriately and if H gets farther from 0 and gets closer to 1 means that the model does not fit the data properly. To have a good model, membership value of real number Y_1 to its estimated fuzzy value $\hat{Y}_1$ must be less than H (Chen Si 2005). Hao and Chiang (2007) achieved more results for fuzzy regression with different kernels.

Application

Data

Data belongs to a gas field located in the Persian Gulf, a region on the boundary of Iran and Qatar. This gas field contains an area around 9700 km². Two main gas bearing formations of this location are Kangan and upper Dalan. The field is constructed of four main independent reservoir layers: K1, K2, K3 and K4. Mineral composition of layers K1 and K3 is from dolomite and anhydrites, layers K2 and K4 composited from limestone and dolomite as well. The carbonate reservoirs in upper Dalan–Kangan formation are heterogenic. For permeability estimation we selected four wells: W1, W2, W3 and W4. We devoted wells W1, W2 and W3 for training purposes and well W4 for testing. In this study we included four well logs providing petrophysical parameters as follows: acoustic-DT, transit interval time or slowness, neutron porosity (NPHI), photoelectric absorption factor (PEF) and gamma ray, intensity of natural radioactivity (GR).

Steps to estimate permeability using MRGC and regression algorithms

- Electrofacies analysis using MRGC
- Applying SVR algorithm on each of the obtained ELs
- Applying FSVR algorithm on each of the obtained ELs
- Correlate predicted permeability obtained from ELs with real permeability from core data

- Merging predicted values for each electrofacies to find permeability log in well W4

Results

Electrofacies analysis using MRGC

First, we applied all four mentioned algorithms to the data; then we compared the results to find the best algorithm.

1. Using MRGC algorithm: we selected 4 and 20 as minimum and maximum number of desired clusters. Then three models of 7, 11 and 13 clusters generated and are used as input for SVR algorithm individually. The results show that determination correlation (DC) of predicted permeability with core data got worse with increasing cluster numbers (DC = 81% for 7, DC = 76% for 11, DC = 66% for 13). So, we chose the model with seven clusters.
2. Using AHC algorithm: we entered different pre-defined number of clusters in AHC algorithm (random numbers of 5, 7 and 11). We obtained clusters with high difference in data distributions. For example, one cluster included only one sample while the other cluster contained 7565 samples. We did not consider it as a proper input for machine learning algorithm because the machine cannot be trained well.
3. Using SOM algorithm: we also tested this algorithm with random numbers of 6, 7 and 11 clusters. This algorithm did not partition the data properly either.
4. Using DC algorithm: we applied this algorithm with randomly pre-defined numbers of 7, 9 and 11. The distribution of data is good while the clusters have more overlap

compared to MRGC. So, we ignored working with this algorithm too.

Finally, we chose MRGC as the best algorithm for our case study to continue the rest of process.

In this work, we had totally 2262 samples. As Fig. 4 shows, first column shows the cluster numbers with their related color. Each color is an EL. Second column is weight numbers which presents number of data points included in each face from total samples. Third column is the average of log value in each face. The reason for numbers in the third column turns back to the material which the formations have made from. For example in the third column, the standard scale for neutron log is between 0.45 and − 0.15. If the third column values are low it is because of the formations such as limestone, dolomite and anhydrite.

Figure 5 displays each well identified with seven colors. By adding segments with the same color, we can create ELs. The graph also shows data points' distribution based on their similarity in each well. For instance, most data points concentrated in facies 7 in well W2 while concentration of samples is mostly in facies 4 in well W3 compared to others. Facies 5 and facies 3 have the most data points in wells W1 and W4, respectively.

Figure 6 shows cross-plots of well logs GR, DT and PEF versus each other including electrofacies data. Geologists and petrophysicists use this information to analyze the area in more detail. Comparison of logs to each other with electrofacies values is very helpful for a better evaluation of the under studied zone. In Fig. 6 clusters might have a slight overlap in some areas. We did not reduce number of clusters to overcome overlapping. First because a slight overlap is normal and second, this zone naturally is made of seven clusters including Limestone, Dolomite-Limestone, Dolostone,

FACIES	Cluster weights	DT (US/F)	GR (GAPI)	NPHI (V/V)	PEF (B/E)
1	292	50.89	29.18	0.04	3.51
2	224	51.74	21.74	0.01	4.80
3	392	50.89	29.18	0.04	3.51
4	420	62.17	18.33	0.04	4.89
5	254	60.51	16.10	0.12	3.51
6	162	52.72	13.95	0.06	3.30
7	518	51.18	16.07	0.02	3.93

Fig. 4 Demonstration of electrofacies information and distribution of total samples in seven clusters. First column shows cluster facies. Second column is number of data points in a cluster. Rest of columns display barycenter of each cluster. Barycenter is the mean of all the points belonging to a cluster
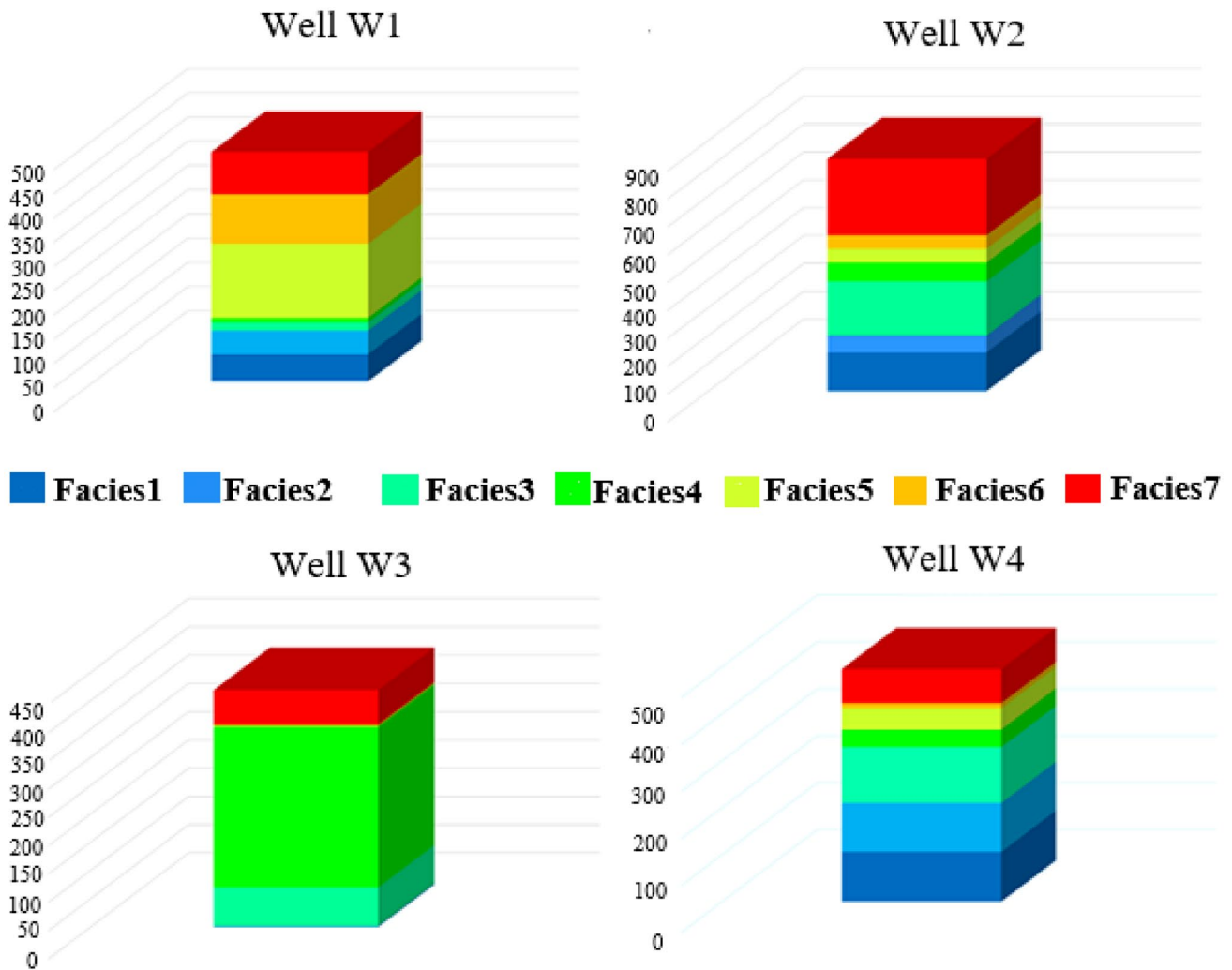
Fig. 5 Demonstration of seven electrofacies recognized by MRGC method for each well in addition to illustration of data distribution in electrofacies. Adding the same color segments from each well makes the ELs

Anhydrite, Evaporite, Shale and Sandstone. Since this clustering is compatible with reality, so we ignored applying any changes.

Calculation of DC for predicted permeability and real data (core plugs) for each electrofacies

In this section, we tried to compare results obtained from petrophysical estimation using two regression methods SVR and FSVR for predicting reservoir properties. Modification of SVR to FSVR helps to reduce noise and outliers existing in initial data. As Table 1 shows, DC calculated from FSVR for each of the partitions (except facies number 3) is remarkably higher than the value obtained from calculated parameter from SVR.

Merging predicted segments to find permeability log in well W4

Figure 7 displays merged parts of estimated permeability in well W4 compared with core data. DC calculated for SVR and FSVR is 0.70 and 0.86, respectively. Obviously, the measured value DC is notably higher for FSVR than SVR.

Permeability prediction using SVR and FSVR algorithms

As previously discussed, we derived seven ELs from well data which are supposed to be employed as the input for SVR and FSVR algorithms in this section. ELs belonging

Fig. 6 Cross-plots of well logs PEF, GR and DT versus each other including electrofacies information for a detailed petrophysical evaluation of the formations
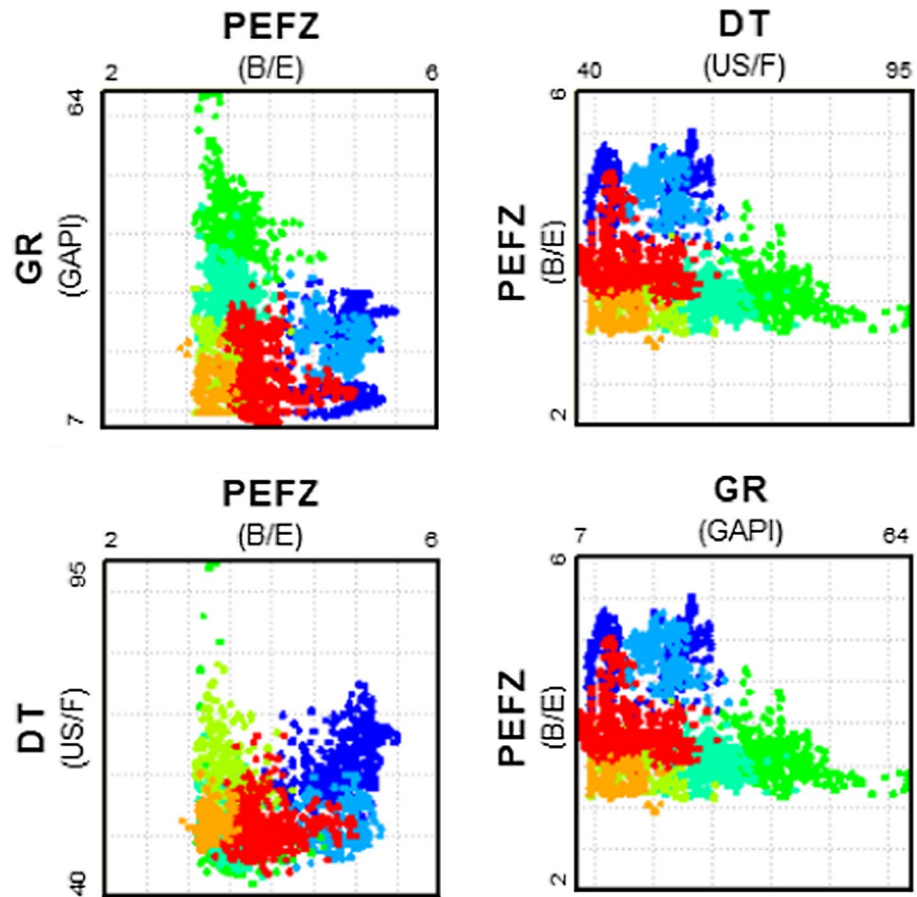

Table 1 DC results calculated for SVR and FSVR for each electrofacies

Facies No	Cluster weights	DC (SVR)	DC (FSVR)
1	292	0.67	0.84
2	224	0.76	0.89
3	392	0.80	0.78
4	420	0.50	0.80
5	254	0.81	0.91
6	162	0.71	0.88
7	518	0.73	0.87

to three first wells: W1, W2 and W3 are considered for training and the electrofacies from well W4 are for testing purpose. We are determined to predict permeability for each electrofacies in well W4 by two algorithms of SVR and FSVR in addition to calculating determination coefficient (DC) of predicted log with real value. For this aim, we first selected one of the electrofacies from well W4, and we examined it for both SVR and FSVR algorithms. Figure 8 shows the estimated permeability for the selected segment which is located in a reservoir zone. Figure 8a shows predicted permeability from SVR and 8b shows the result obtained from FSVR. Amount of DC is 0.69 and 0.87 for SVR and FSVR, respectively. Figure 9 shows the scatter plot of predicted permeability versus real permeability (permeability measured from core plugs) for each electrofacies in well W4. First column displays the color of each facies, second column relates to SVR results and the third column devotes to FSVR. Generally, what is notable from scatter plots of FSVR column, data points are more concentrated along the regression line.
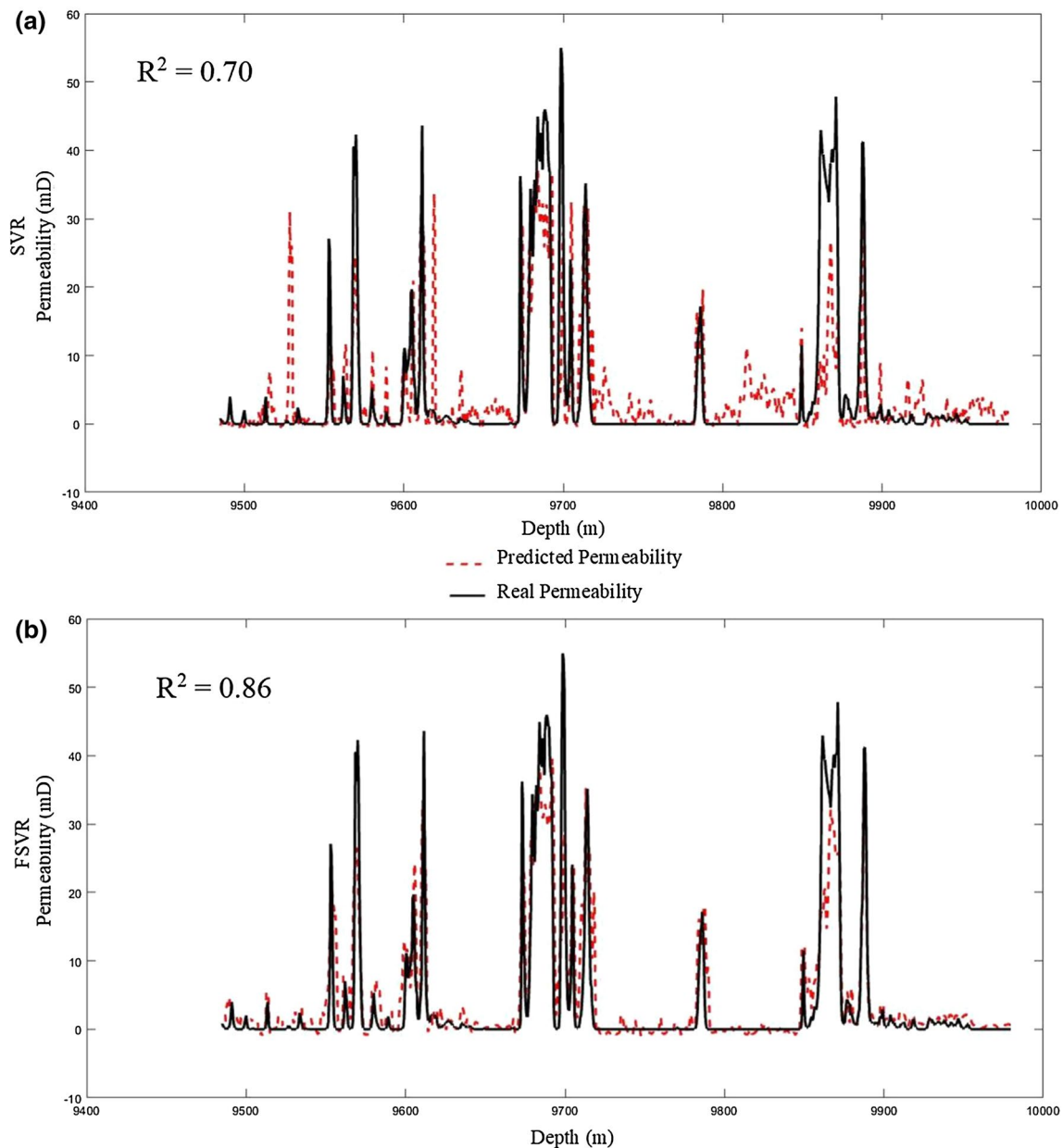
Fig. 7 Merged parts of estimated permeability for each electrofacies in well W4 compared with real permeability (core data obtained from core plugs in laboratory). Predicted permeability showed with dotted red line and real permeability displayed with black line. DC calculated for **a** SVR and **b** FSVR is 0.70 and 0.86, respectively

Conclusion

In this study we took advantage of EA using MRGC method to predict permeability utilizing well logs of a gas field located in Persian Gulf. There are some points while estimating permeability. Firstly, permeability is a very complicated parameter which changes dramatically in vertical scales.

Secondly, this parameter has a tightly relationship with formation structure. Thirdly, studies of reservoir zones with core missing data via regression methods is almost impossible because core data is needed for validation operations in such algorithms. According to all these three facts, we took advantage of EA to overcome these obstacles for modeling reservoir petrophysical parameters. The results acquired in
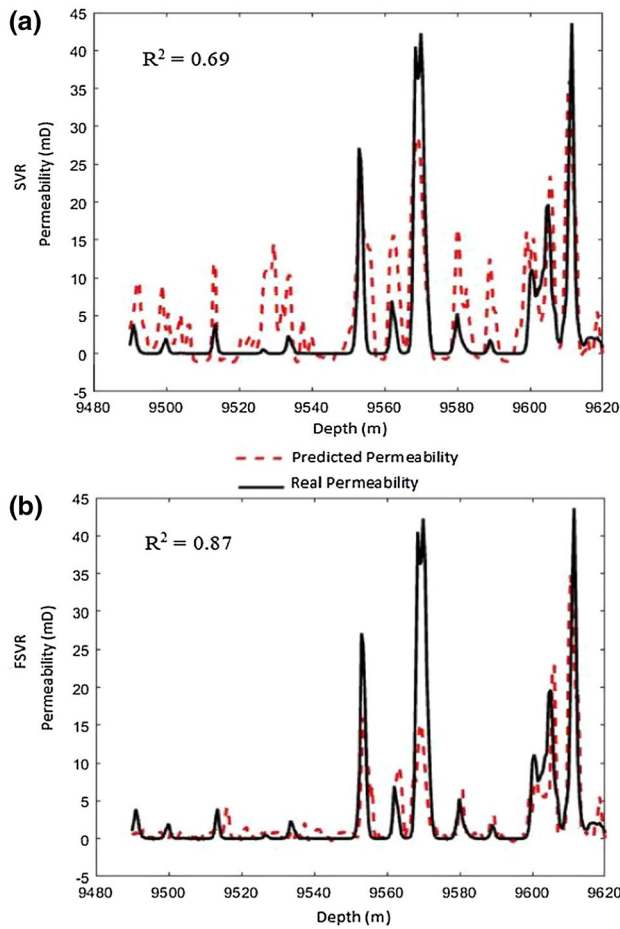
Fig. 8 Estimated permeability for one of the electrofacies belonging to a reservoir zone in well W4. **a** SVR, **b** FSVR. Predicted plot is displayed with dotted red line and real permeability (permeability obtained from core plugs in the laboratory) is showed with black line. Calculated DC for SVR and FSVR is 0.74 and 0.83, respectively

our paper shows the applicability of EA and MRGC method to detect variety of strata at well locations which was so helpful to solve abrupt changes of permeability problem and use it for well locations with core missing data. Furthermore, modification of SVR to FSVR proved to be effective to suppress noise and outliers as inevitable part of geophysical data. Consequently, the results show estimated permeability using FSVR is superior to SVR algorithm. Calculated determination coefficients prove this fact by providing us with notably higher values for FSVR compared to SVR.

Fig. 9 Scatter plots of predicted permeability versus real permeability (permeability obtained from core plugs in the laboratory) for each electrofacies in well W4. First column shows facies with their related color, second and third columns are results from SVR and FSVR, respectively ▶
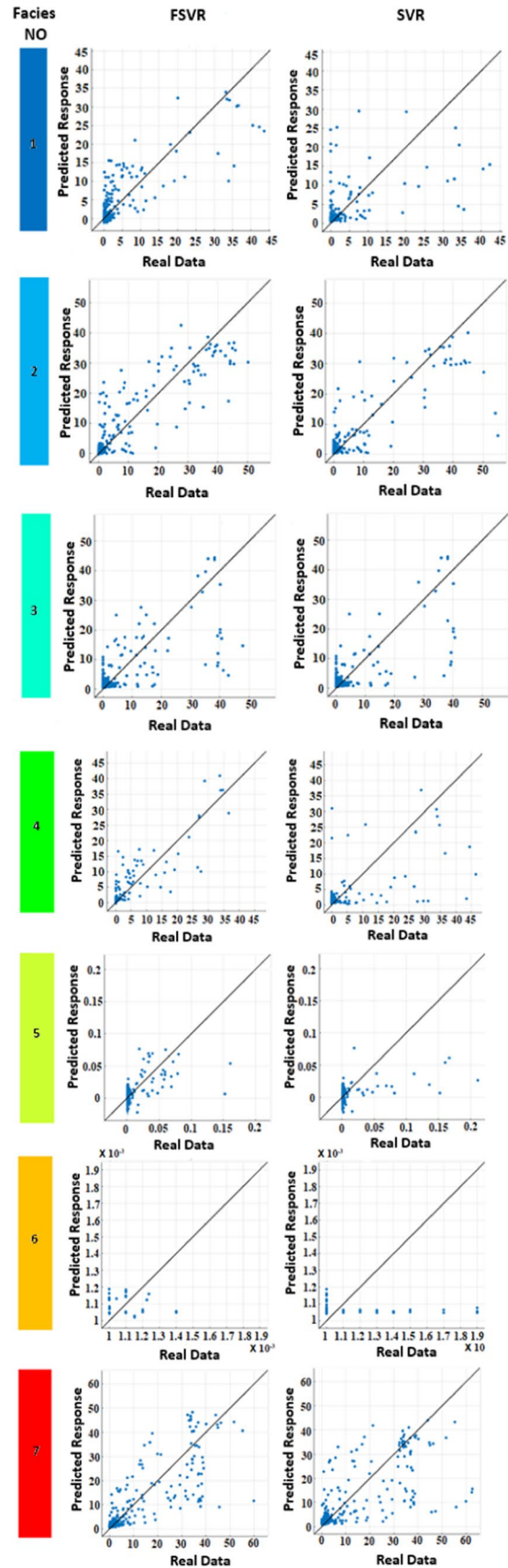
Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Aghchelou M, Nabi-Bidhendi M, Shahvar M.B (2012) Lithofacies estimation by multi resolution graph-based clustering of petrophysical well logs: case study of south pars gas field of Iran, SPE 16299

Al-Anazi AF, Gates DI (2012) Support vector regression to predict porosity and permeability: effect of sample size. Comp Geosci 39:64

Al-Mudhafar W, (2020) Integrating electrofacies and well logging data into regression and machine learning approaches for improved permeability estimation in a carbonate reservoir in a giant southern iraqi oil field, offshore technology conference, houston, Texas, ISBN: 97–1–61399–707–9

Ameri S, Molnar D and Mohaghegh S (1993) Permeability evaluation in heterogeneous formations using geophysical well logs and geological interpretations SPE Western Regional Meeting (Anchorage, AK) SPE 26060

Bishop CM (2006) Pattern recognition and machine learning. Springer, New York, pp 325–344

Chen Si B (2005) Determining soil hydraulic properties from tension infiltrometer measurements: fuzzy regression. Soil Sci Soc Am J 69:1931–1941

Criollo D., Marin Z, Vasquez D (2016) Advanced electrofacies modelling and permeability prediction: a case study incorporating multi-resolution core, nmr and image log textural information into a carbonate facies study, SPWLA 22nd Formation evaluation symposium of Japan, SPWLA-JFES-2016-K

Davis J (2018) Handbook of mathematical geosciences, In: Electrofacies in Reservoir Characterization, Springer, Cham, pp 211–222.

Fiandaca G, Maurya PK, Balbarini N, Hordt A, Christiansen AV, Foged N, Bjerg PL, Auken E (2018) Permeability estimation directly from logging-while-drilling induced polarization data. Water Resour Res 54(4):2851–2870

Gan Z, Griffin T, Dacy J, Xie H., Lee R, (2018) Fast pressure-decay core permeability measurement for tight rocks, SPWLA 59th Annual logging symposium, London, UK.

Glover WJP (2000) Petrophysics MSc Course Notes. Department of geology and petroleum geology university of Aberdeen, UK, pp 21–31

Gunn SR (1998) Support vector machines for classification and regression, University of Southampton, 10 Ma.

Hao P-Y, Chiang J-H (2007) A fuzzy model of support vector regression machine. Int J Fuzzy Syst 9(1):45–50

Hasan M, Sobhan M (2020) Describing fuzzy membership function and detecting the outlier by using five number summary of data. Am J Comp Math 10:410–424

Hosseinzadeh S, Shaghaghi M (2020) GPR data regression and clustering by the fuzzy support vector machine and regression. Progress in Electromagn Res M 93:175–184

Iturraran-Viveros U, Parra J (2014) Artificial neural networks applied to estimate permeability, porosity and intrinsic attenuation using seismic attributes and well-log data. J Appl Geophys 107:45–54

Khan Mohammadi M, Sherkati SH (2010) Fracturing analysis in south pars gas feild. Explor Prod Mon 77:43–49

Khoshbakht F, Mohammadnia M (2013) Assessment of clustering methods for predicting permeability in a heterogeneous carbonate reservoir, J Pet Sci Technol, 2(2)

Klinkenberg LJ (1941) The permeability of porous media to liquids and gases. In: Drilling and production practice. American Petroleum Institute

Kozeny J (1927) Uber kapillare leitung der wasser in boden. royal academy of science, Vienna. Proc Class I 136:271–306

Le VH, Liu F, Tran DK (2009) Fuzzy linguistic logic programming and its applications. Theory Practice Logic Progr. https://doi.org/10.1017/S1471068409003779

Leverett MC (1941) Capillary behaviour in porous solids Paper SPE-941152-G. AIME Trans. https://doi.org/10.2118/941152-G

Li Z, Xie Y, Li X, Zhao W (2021) Prediction and application of porosity based on support vector regression model optimized by adaptive dragonfly algorithm. Energy Sour, Part A: Recover, Util Environ Eff. https://doi.org/10.1080/15567036.2019.1634775

Liaghat M, Nuraei-nedhad M, Adabi M (2021) Determination and interpretation of electrofacies using som neural network and its application to prediction of khami group facies in marun oil field (South West Iran). J Pet Res 31:96–111

Lim KM, Sim YC, Oh KW (2002) A face recognition system using fuzzy logic and artificial neural network, IEEE. [1992 Proceedings] IEEE International conference on fuzzy systems, USA

Lin C, Wang S (2002) Fuzzy support vector machines. IEEE Trans Net. https://doi.org/10.1109/72991432

Mohaghegh S (2000) Virtual-intelligence applications in petroleum engineering: part I Artificial neural networks. J Pet Technol 52:64–73

Nashawi IS, Malallah A (2009) Improved electrofacies characterization and permeability predictions in sandstone reservoirs using a data mining and expert system approach. Petrophysics 50(03)

Negi JK, Verma CP, Kumar A, Prasad US, Lal C (2006) Predicting lithofacies using artificial neural network and log-core correlations, 6th international conference and exposition on petroleum geophysics, Kolkata. pp 809–811

Nourani V, Alami MT, Vousoughi FD (2016) Self-organizing map clustering technique for ANN- based spatiotemporal modeling of groundwater quality parameters. J Hydroinf 18:288–309

Novikoff ABJ (1962) On convergence proofs on perceptrons. In: Proceedings of the symposium on the mathematical theory of automata, 12: 615–622

Okwu M, Nwachukwu AN (2019) A review of fuzzy logic applications in petroleum exploration, production and distribution operations. J Petrol Explor Prod Technol 9:1555–1568

Rafik B, Kamel B (2016) Prediction of permeability and porosity from well log data using the nonparametric regression with multivariate analysis and neural network Hassi R'mel Field, Algeria. Egyptian J Pet 26(3):763–778

Rustam Z, Hidayat, Nurrimah R (2019) Indonesia composite index prediction using fuzzy support vector regression with fisher score feature selection. Int J Adv Sci Eng Inform Technol 9(1):21

Serra O, Abbott HT (1980) The contribution of logging data to sedimentology and stratigraphy. In: SPE 9270, 55th technical conference, Dallas, TX, pp 19

Shen S, Fang Z, Li X (2020) (2020), laboratory measurements of the relative permeability of coal, a review. Energies 13(21):5568

Silva L, Avanis G, Schiozer D (2020) Support vector regression for petroleum reservoir production forecast considering geostatistical realizations. SPE Res Eval Eng 23(04):1343–1357

Singh H, Myshakin E, Seol Y (2020) A novel relative permeability model for gas and water flow in hydrate-bearing sediments with laboratory and field-scale application. Sci Rep. https://doi.org/10.1038/s41598-020-62284-5

Smith C, Hamilton L (2014) Carbonate reservoir permeability from nuclear magnetic resonance logsIPTC-17869-MS Int. Petroleum Technology Conf. Kuala Lumpur

Suykens JAK, Van Gestel T, Brabanter J, De Moor B, Vandewalle J (2002) Least squares support vector machines. World Scientific, Singapore

Thomeer JH (1960) Introduction of a pore geometrical factor defined by the capillary pressure curve. J Pet Technol. https://doi.org/10.2118/1324-G

Vapnik VN (1982) Estimation of dependences based on empirical data. Springer, Berlin

Vapnik VN (1995) The nature of statitical learning theory. Springer, New York

Vapnik VN, Chervonenkis A (1979) Theory of pattern recognition [in Russian]. Nauka, Moscow (German Translation: Wapnik W, Tscherwonenkis A, Theories der Zeichenerkennung, Akademie-Verlag, Berlin

Ye S, Rabiller P, (2000) A new tool for electrofacies analysis: multi resolution graph-based clustering, SPWLA 41st annual logging symposiu.

Zadeh. L.A. (1965) Fuzzy Sets. Inf Control 8(3):338–353

Zhang D, Lin J, Peng Q, Wang D, Yang T, Sorooshian S, Liu X, Zhuang J (2018) Modeling and simulating of reservoir operation using the artificial neural network support vector regression, deep learning algorithm. J Hydrol 565:720–736

Zhang R, Duan X, Hao L, (2010) Fuzzy support vector regression for function approximation with noises, International conference on computer application and system modeling (ICCASM 2010)