



ISSN: 2785-2997

Journal of Human, Earth, and Future

Vol. 2, No. 2, June, 2021



A Comparison of Machine Learning Approaches for Prediction of Permeability using Well Log Data in the Hydrocarbon Reservoirs

Mohsen Talebkeikhah ¹, Zahra Sadeghtabaghi ^{2*}, Mehdi Shabani ²

¹ Department of Civil Engineering, École Polytechnique Fédérale de Lausanne EPFL, CH-1015 Lausanne, Switzerland

² Department of Petroleum Engineering, Amirkabir University of Technology, Tehran, Iran

Received 09 March 2021; Revised 16 May 2021; Accepted 25 May 2021; Published 01 June 2021

Abstract

Permeability is a vital parameter in reservoir engineering that affects production directly. Since this parameter's significance is obvious, finding a way to an accurate determination of permeability is essential as well. In this paper, the permeability of two notable carbonate reservoirs (Ilam and Sarvak) in the southwest of Iran was predicted by several different methods, and the level of accuracy in all models was compared. For this purpose, Multi-Layer Perceptron Neural Network (MLP), Radial Basis Function Neural Network (RBF), Support Vector Regression (SVR), decision tree (DT), and random forest (RF) methods were chosen. The full set of real well-logging data was investigated by random forest, and five of them were selected as the potent variables. At depth, computed gamma-ray log (CGR), spectral gamma-ray log (SGR), neutron porosity log (NPHI), and density log (RHOB) were considered efficacious variables and used as input data, while permeability was considered output. It should be noted that permeability values are derived from core analysis. Statistical parameters like the coefficient of determination (R^2), root mean square error (RMSE), and standard deviation (SD) were determined for the train, test, and total sets. Based on statistical and graphical results, the SVM and DT models perform more accurately than others. RMSE, SD and R^2 values of SVM and DT models are 0.38, 1.63, 0.97 and 0.44, 2.89, and 0.96 respectively. The results of the best-proposed models in this paper were then compared with the outcome of the empirical equation for permeability prediction. The comparison indicates that artificial intelligence methods perform more accurately than traditional methods for permeability estimation, such as proposed equations.

Keywords: Permeability; Well Logging; Machine Learning; Data Mining; Ilam Formation; Sarvak Formation.

1. Introduction

Permeability can be considered as one of the most significant concepts in reservoir engineering owing to its effect on identifying flow units, characterization of reservoirs, and planning for perforation [1-3]. Many challenges faced during production could be related to this parameter. Moreover, permeability is required for the dynamic modeling of a reservoir. Therefore, accurately determining this parameter is of considerable importance. There are several common methods for this target, such as well testing, core analysis, and well logging. Usually, because of high cost and time consumption, these methods are not available in all wells of a field or all over the desired intervals, and sometimes the low accuracy of some methods causes us to avoid using their results. Thus finding a model that predicts permeability values of a reservoir can provide an insight to act better in so many branches like developing a plan of field, production, reserve estimation, etc. This paper aims to predict the permeability's precise values in two Iranian significant carbonate reservoirs, the Ilam and Sarvak Formations.

* Corresponding author: zahrasadeghtabaghi@aut.ac.ir

<http://dx.doi.org/10.28991/HEF-2021-02-02-01>

➤ This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

© Authors retain all copyrights.

Artificial intelligence and machine learning are considered widely used techniques for parameter prediction, especially in different branches of petroleum engineering. For instance, these artificial intelligence methods have been used for making synthetic well logs [4], determination of infill well location [5], reservoir characterization [6], drilling fluid [7], oil gas ratio [8], prediction of permeability [9], lost circulation [10], prediction of scale thickness [11], and vitrinite reflectance prediction [12]. Various approaches can predict this parameter by making models, and many published studies have investigated the subject of permeability modeling. In some case studies [1, 13-16], neural networks were used to predict permeability values while well-logging data was used as input. Other efforts have been made by applying the committee machine neural network for this goal [17, 18]. In 2014, two RVR and SVR methods were applied for permeability prediction through well log data [19]. RVR or Relevance Vector Regression is one of the algorithms of artificial intelligence that was compared with SVR or Support Vector Regression. Final results indicated that RVR is more precise for permeability prediction. Saljooghi & Hezarkhani (2015) used a method based on wavelet theory and artificial neural networks for the prediction of permeability [20]. The prediction of NMR logging parameters was made by Jamshidian et al. (2015) [21]. For this purpose, they utilized a multilinear perceptron network optimized by the imperialist competitive algorithm. Handhal (2016) predicted permeability through porosity values and found a promising relationship between these two parameters by using approaches like M5 decision trees, linear regression, and adaptive neuro-fuzzy inference system [22]. Rafik & Kamel (2017) proposed a new procedure on this issue and predicted permeability with the combination method of a stepwise algorithm and alternating conditional expectations [23]. For example, Al-Amri et al. (2017) and Zhao et al. (2018) used multi-resolution graph-based clustering (MRGC) and representative elementary volume (REV) scale lattice Boltzmann method (LBM) for permeability prediction, respectively [24, 25]. Eventually, Wood (2020) used the transparent open box machine-learning and data-mining algorithm for this aim [26]. Already, the Sarvak Formation was studied [27–29], whereas the Ilam Formation was not investigated. This paper peruses these two reservoirs for the first time in the studied field.

In this paper, permeability quantities are predicted with high precision for Ilam and Sarvak reservoirs, which are two significant reservoir intervals in Iranian oil fields. A full set of well logging data and core measurements are available for these formations in a well in the southwestern part of Iran. Artificial neural networks such as Multi-Layer Perceptron Neural Network (MLP), Radial Basis Function Neural Network (RBF), and Support Vector Regression (SVR) are used for predicting permeability. In addition to these approaches, two data mining methods, decision tree (DT) and random forest (RF), are applied for permeability prediction. Finally, all the techniques' results are compared by known statistical parameters to determine the most accurate model. A comparison of all these methods for finding permeability values in the studied field is the primary goal of this paper, whereas previous studies have not focused on such a comprehensive comparison, and they were summarized in just one or two distinct methods. This point can be considered the main innovation of this paper. Also, there is no study about the permeability prediction of the Ilam and Sarvak formations. Moreover, a comparison between these artificial intelligence techniques and the proposed equations for permeability prediction in Ilam and Sarvak was made.

2. Geographical, Geological, and Data Descriptions

The southwestern part of Iran, with great amounts of hydrocarbon reserves, is considered one of the world's most productive regions [30]. The studied well is located in the Abadan plain, which is a part of this area (Figure 1). With an approximate area of 26000 km², the Abadan plain is located between the boundary of Iran and Iraq, Persian Gulf, and Dezful Embayment that containing numerous oil and gas wells [31]. Jurassic and Cretaceous reservoirs of this region include lots of oil and gas [32–34]. Ilam, Sarvak, Gadvan, and Fahliyan are considered oil reservoirs in this region. Still, just limy Ilam and Sarvak formations with an approximate thickness of 90 and 600 m, respectively, are investigated in this paper.

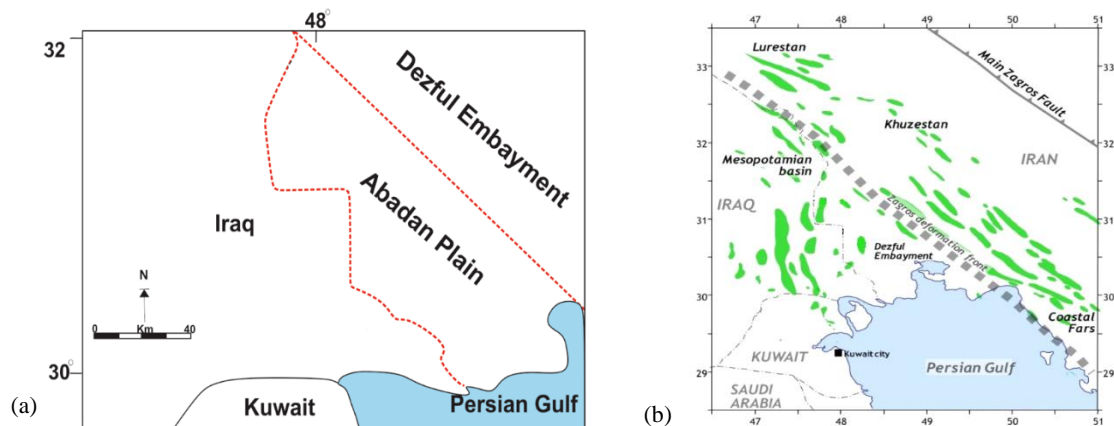


Figure 1. Location of Abadan Plain (a) and oil fields in SW Iran shown with green color (b) after [33, 34]

The Santonian Ilam Formation, a proven significant hydrocarbon reservoir interval, comprises light gray shallow-marine limestones with beds of algal and rudist-bearing limestones capped by shales. These shallow-water carbonates contain natural fractures [35]. The late Albian–early Turonian Sarvak Formation is the most critical reservoir in the Abadan plain. The Sarvak Formation is composed of a shallow-water massive limestone and deep-marine thin-bedded facies [36]. There are two essential unconformity surfaces in Cenomanian-Turonian intervals of the Sarvak Formation in this area that control the reservoir's quality. In the Abadan Plain, an argillaceous-shaley limestone (Laffan Formation with the age of Coniacian) with a thickness of about 10-15 m separates the Sarvak and Ilam formations [37].

Previous research indicates the Ilam reservoir is considered a diagenetic reservoir in Dezful embayment and possesses five units that one of them is an excellent reservoir with high quality [38]. Sarvak Formation keeps the best quality of a reservoir in intervals with rudist limestones over the Zagros area and the Persian Gulf [39]. Also, Ilam and Sarvak are fractured reservoirs in the Strait of Hormuz, the Persian Gulf [40]. The other study about these two reservoirs mentioned earlier indicates that Ilam Formation shows better reservoir characteristics than Sarvak in the Dezful embayment [41].

In this work, full set logging data of computed gamma ray (CGR), spectral gamma ray (SGR), potassium (POTA), thorium (TH), uranium (URAN), sonic (DT), Density (RHOB), Neutron (NPHI), photoelectric absorption properties (PEF), Micro-Spherical Focused Log (MSFL), laterolog deep (LLD), and laterolog shallow (LLS) belong to an oil well located in the Abadan plain. The statistical details of these data have been summarized in Table 1. Moreover, the core samples of Ilam and Sarvak formations were analyzed in the core laboratory. The results of measured permeability are also shown in Table 1. Different computational intelligence methods have been used to model the permeability employing a dataset containing 171 data points.

Table 1. Input and output data with the minimum, maximum, and average values

#	Parameter	Unit	Min	Max	Average
1	CGR	API	0.58	54.98	15.11
2	DT	US/F	47.50	81.49	62.63
3	LLD	OHMM	0.46	3074.19	33.95
4	LLS	OHMM	0.45	3073.80	29.86
5	MSFL	OHMM	0.10	175.03	8.33
6	NPHI	V/V	0.0021	0.27	0.10
7	PEF	B/E	3.48	11.26	5.95
8	POTA	%	0.03	2.93	0.91
9	RHOB	G/C ³	2.24	2.78	2.54
10	SGR	API	13.79	89.45	28.12
11	TH	PPM	0.03	5.75	0.68
12	URAN	PPM	0.06	7.08	1.45
13	Depth	M	2670	3430	-
14	Permeability (K)	mD	0.01	817.12	14.65

After gathering data, the required data were selected and then normalized to enhance the system's accuracy [42]. The dataset is complex and constructed as a model in machine learning due to working with many features. So, selecting essential elements and the decreasing dimension of the dataset seems necessary [43]. Also, this causes reducing the time of training, makes the interpretation of the model easier, and decreases expenses of data gathering [44]. In this technique, variables' choice and importance depend on the selected predictive model [45]. Thus, for finding variables with the most impact, an appropriate model should be employed. In this paper, well logging data are considered as inputs. The variation of all these values with depth is shown in Figure 2. The random forest approach was used to determine the importance of variables, which will be discussed later. The results of the random forest method are illustrated in Figure 3. As shown, the first and the second most effective variables on permeability are depth and NPHI followed by RHOB, CGR, and SGR. Therefore, Depth, NPHI, RHOB, CGR, and SGR were selected for making models as important inputs with the most impact. It must be emphasized that this data selection is suitable for this situation and cannot be attributed to all reservoirs. It should be noted that Figure 4 demonstrates all possible relations between these selected variables and permeability. As indicated in this figure, the value of the regression coefficient of CGR and NPHI versus permeability is negative, while for two other inputs, it has a positive correlation.

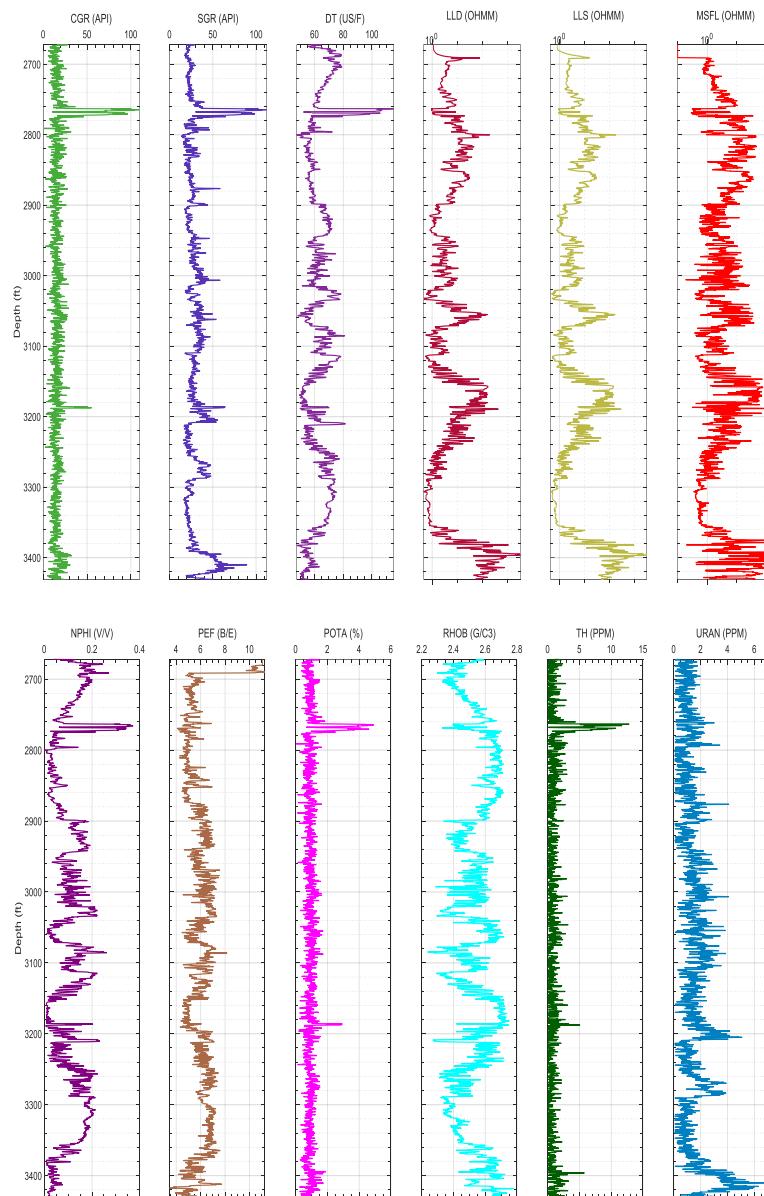


Figure 2. Full set well logging data

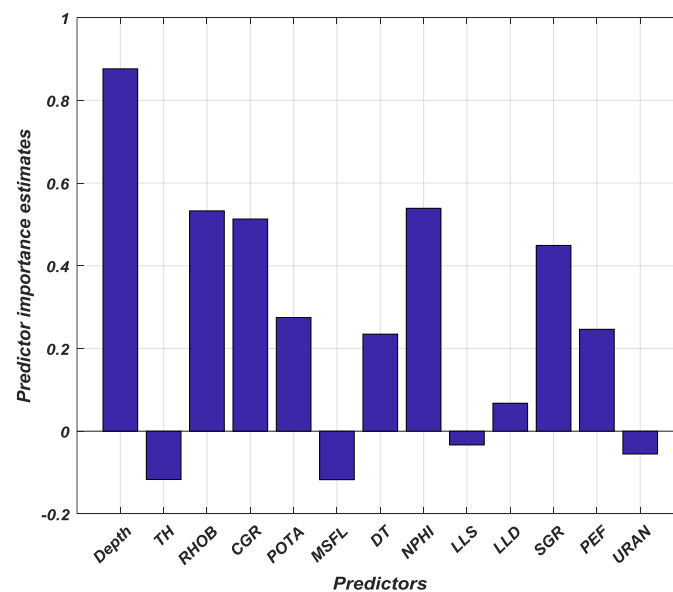


Figure 3. Feature ranking by random forest

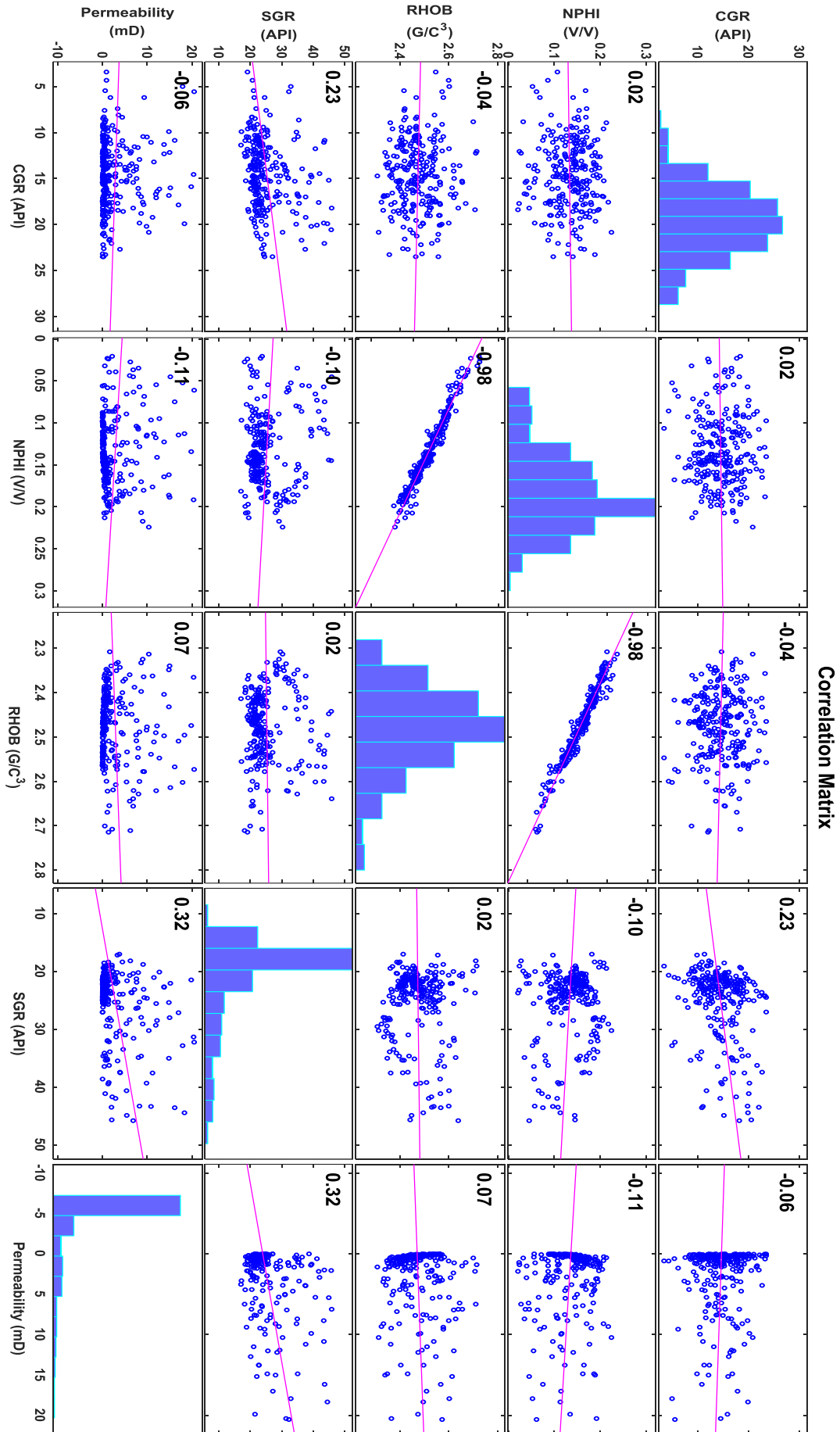


Figure 4. Charts of regression coefficient between selected input and permeability

Application of RF method as a set of some unpruned decision trees with sequential growth in place of a single restricted type was mentioned by Breiman (2001) at first. RF uses the bootstrap sampling method for choosing random data due to substitution eventuality. Then, the other samples are used for the test of the tree. All trees will have this procedure, so the estimation will act better since random sampling makes differences between sets of trees. Eventually, with an internal cross-validation program between all trees, the average of errors is calculated. For training a tree in this approach, a small subset of size m out of M features for each node is chosen [50]. While altering features for each node, all over forest growth, the size of m is sustained stable. So, the best splitting of nodes is done in m features, not all features of M [51]. Different parameters in three categories of trees in the forest (ntree), size of subset (mtry), and min leaf size in each tree grow up the trees in RF. Indeed, the optimum value of these parameters leads to convincing prediction between ensembles of trees. By replacing features randomly, the precision of the model is calculated. Thus, the rank of features regarding their effects on the outputs of model is determined and can be considered an advantage for RF.

Numerous errors, the great value of variance, and the difficulty of overfitting are the problems of Decision trees; however, they can act adequately in nonlinear systems [52]. It must be emphasized that these aforementioned issues can be avoided by applying random forest, an ensemble learning method suggested by Breiman et al. (1984) [53]. The random forest can be considered an appropriate tool for unsupervised learning and classification and regression [54]. Similar to other ensemble learning techniques, the random forest can improve the efficiency of some weak learners, such as single decision trees, single perception, etc. [55].

The structure of ANNs was derived from the brain's structure, and they simulate the function of the brain for learning [56]. Since ANNs are mathematically non-linear, available, and simple models, they are used extremely [57, 58]. One of the methods of artificial intelligence is the multi-layer perceptron network used widely [59, 60]. These networks act appropriately in non-linear systems modeling. Hidden layer(s) provide the relation of input and output layers in a model. The effectiveness of an MLP-NN strongly depends on the number of hidden layers and the number of their neurons [61, 62]. It should be noted that even one hidden layer has high efficiency for constructing a proper model, and networks with two or more hidden layers are appropriate for systems that are so complicated. Multiplying the weight of every neuron in related neurons in the previous layer and summing them eventuate the value of neurons in the hidden layer or output layer. Then a bias is added, and an activation function is exerted on this value to determine the final output. There are some kinds of the activation function. Usually, tnsig is used for the hidden layer and purelin for the output layer [63, 64], which are as follows:

$$\text{Purelin: } f(x) = x \quad (1)$$

$$\text{Tansig: } f(x) = \frac{2}{1 + \exp(-2x)} - 1 \quad (2)$$

Also, the Levenberg Marquardt algorithm's performance in estimating efficiency function between the other mentioned learning algorithms is the best. In this paper, this algorithm is used. To determine the most suitable network with the high ability for modeling non-linear problems, the tansig function for the hidden layer and purelin for the output layer were selected. The numbers of neurons were trained, and through a trial and error procedure, the best performance was selected by comparison. Meanwhile, MSE or mean squared error was calculated in all iterations during training. Figure 6 displays MSE values for different numbers of neurons in both hidden layers. According to the figure, MLP structure with two hidden layers in which the first one has three neurons and the second one has five neurons is the best one with minimum MSE. All characteristics of the network are presented in Table 2.

Table 2. Structure of the MLP network developed to predict permeability

Network type	Multi-Layer Perceptron
Training function	Levenberg-Marquardt backpropagation
Number of layers	3
Number of neurons in 1st hidden layer	3
Number of neurons in 2nd hidden layer	5
Transfer function of 1st hidden layer	TANSIG
Transfer function of 2nd hidden layer	TANSIG
Number of neurons in output layer	1
Transfer function of output layer	PURELIN
Performance objective function	MSE

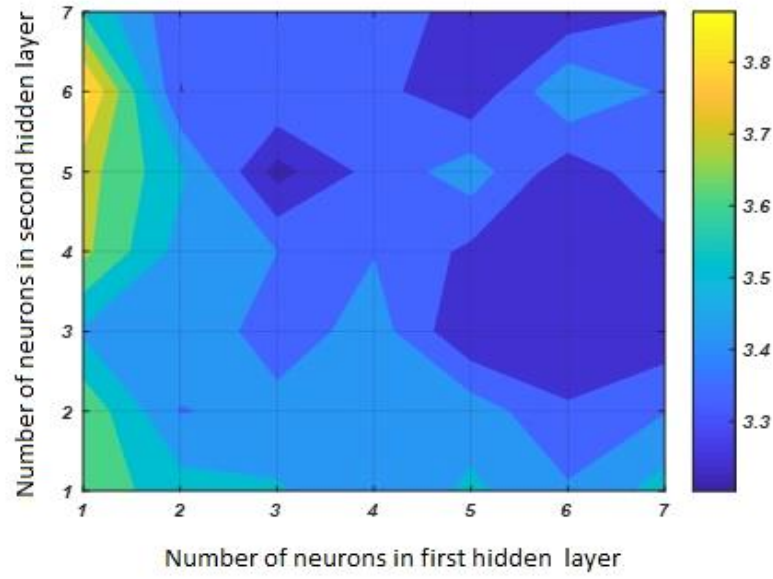


Figure 6. The number of neurons of hidden layers versus MSE of the training data set

Since RBFs behave more correctly in non-linear system modeling, they are used routinely [65]. To model nonlinear data, these networks have more proper operations for better accuracy. The other benefit of them is their single procedure for finding the answer in contrast to MLP networks having iterative procedure [66]. There are different kinds of methods for finding the center of RBF networks, such as random selection of centers, clustering, and density estimation. In this paper, clustering has been used. The precision of K-means is better than the other methods for finding centers, therefore it is used more for clustering. Below equation shows the output of RBF [67, 68]:

$$f(x_i) = \sum_{i=1}^N w_i \varphi_{ki}(\|x_k - c_i\|) \quad i = 1, \dots, N \quad k = 1, \dots, M \quad (3)$$

where, c_i is the center, N is the number of hidden layers' neurons, $\|x_k - c_i\|$ is the distance of inputs and centers, w_i is the weight parameter, φ_{ki} is the kernel function, and M is number of inputs.

The schematic form of the RBF network, which is used in this paper, is shown in Figure 7. In addition, determining the proper number of neurons in the hidden layer is done by the RBF network. For this purpose, the number of neurons has been added one by one, and the MSE value is calculated. The minimum amount of MSE corresponds to the best structure of the network. As illustrated in Figure 8, a network with 25 neurons can perform in the best condition. Moreover, SVR is a technique that can estimate the purposed variables through statistical learning theory principals. By using an appropriate linear regression model in a new feature space that input data which are mapped from the original space to go to a higher m-dimensional space, unknown values are estimated. Different phases of this technique are displayed in Figure 9. As a final point, it should be stated that in this paper, two main ways have been used for determining optimum parameters of all implemented artificial intelligence approaches. The procedure of previous research [10], as well as the trial and error approach, have been used for this purpose.

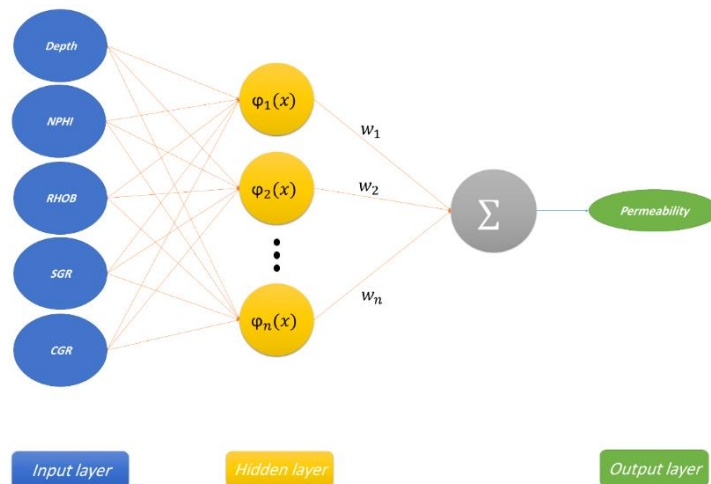


Figure 7. Structure of the RBF network with three layers.

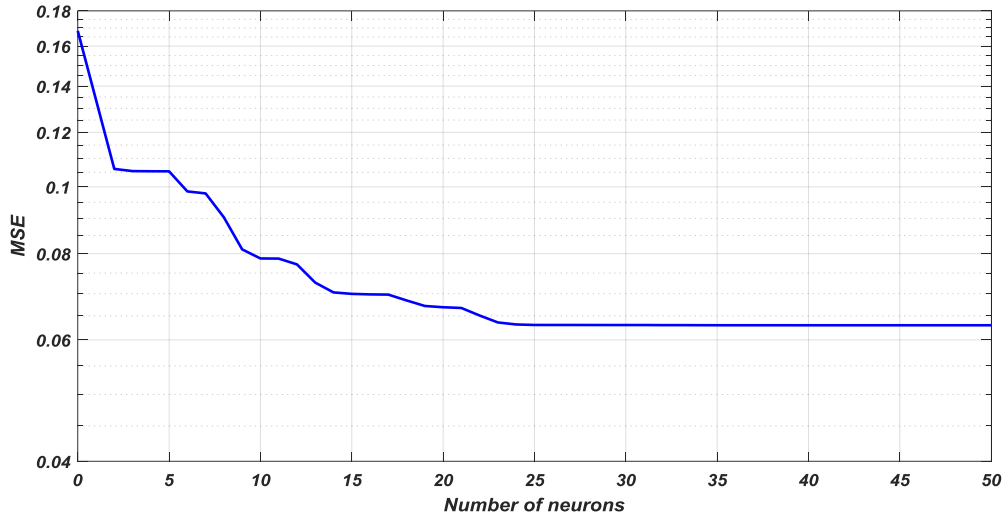


Figure 8. MSE versus the number of hidden layer's neurons for determining the best structure of the network

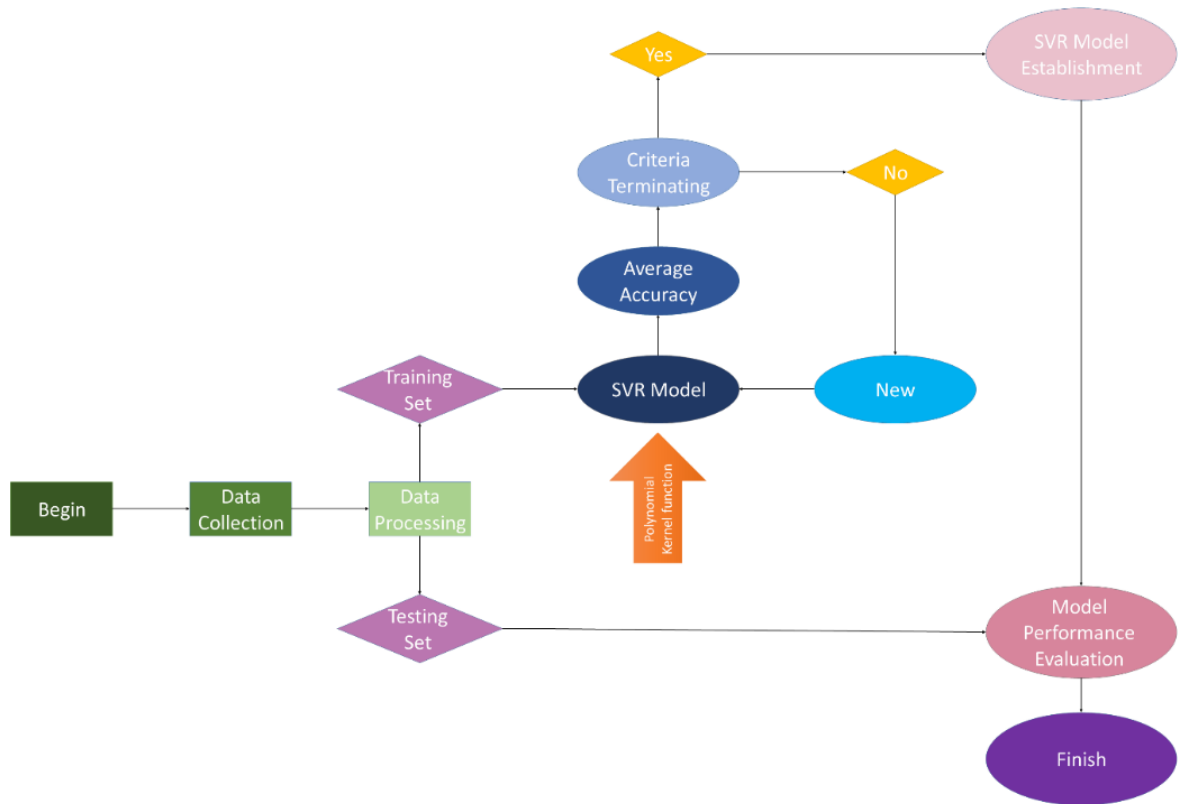


Figure 9. The sequence of steps in the SVR model

4. Results and Discussion

This section of the paper focuses on analyzing models to select the best of them. Some statistical and graphical methods are used for this purpose [43, 69].

It should be noted that statistical parameters should investigate the performance and precision of this paper's implemented models. Therefore, analysis of errors is done with some indices named Coefficient of determination (R^2), Root mean square error (RMSE) and Standard deviation (SD), which are defined respectively as bellows [70]:

$$R^2 = 1 - \frac{\sum_{x=1}^m (y_x - z_x)^2}{\sum_{x=1}^m (y_x - y_{x,mean})^2} \quad (4)$$

$$RSME = \sqrt{\frac{1}{m} \sum_{x=1}^m (y_x - z_x)^2} \quad (5)$$

$$SD = \sqrt{\frac{1}{m-1} \sum_{x=1}^m \left(\frac{y_x - z_x}{y_x} \right)^2} \quad (6)$$

where, y_x is experimental value, z_x is estimated value and $y_{x,mean}$ is the average of experimental values.

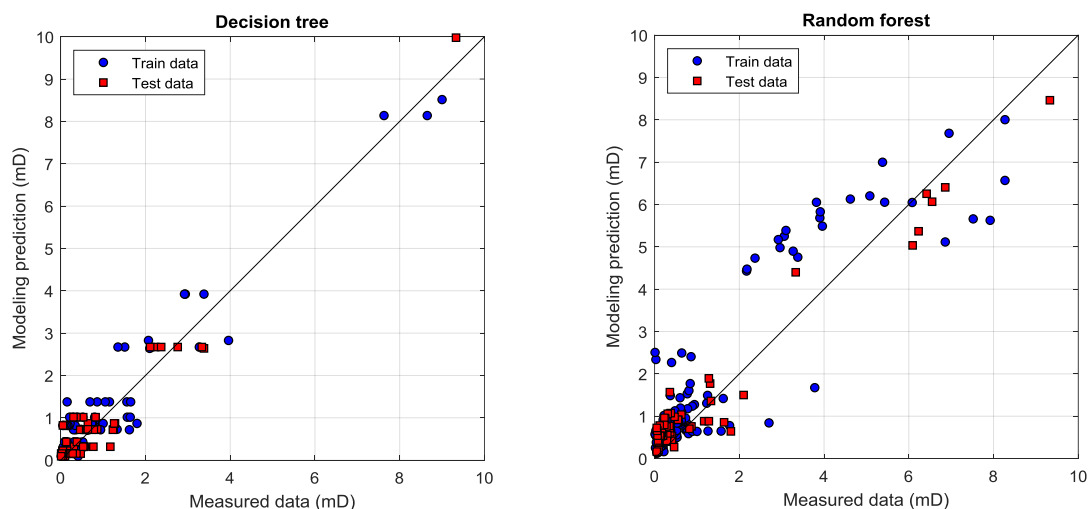
RMSE determines dispersion of data around the regression line whereas R^2 indicates the matching between initial and predicted data. Besides, standard deviation or SD is a measure of a set of values variation. Therefore, these parameters can evaluate implemented models' precision by finding the discrepancies between measured data and predicted data.

These indices are calculated and presented in Table 3, revealing high accuracy of models, especially DT and SVM, and less precision for the RF model. As listed in this table, statistical parameters have been calculated for training, test, and total data sets. The numbers of data in training and test sets are 120 and 51, respectively. In the total set, the highest R^2 values belong to the SVM and DT methods (0.97 and 0.96). As shown in Table 3, SVM and DT models possess the lowest amount of RMSE and SD. Numerical assessments indicate the more precision of these two methods, whereas the other implemented models like RF, RBF, and MLP also show relatively high accuracy.

Table 3. Statistical parameter values for SVM, DT, RF, RBF, and MLP

Statistical parameters	SVM	DT	RF	RBF	MLP
Training Set					
RMSE	0.24	0.46	0.79	0.63	0.45
SD	1.76	2.64	8.85	5.71	5.77
R^2	0.99	0.96	0.88	0.94	0.96
Number of Data	120	120	120	120	120
Test Set					
RMSE	0.58	0.39	0.55	0.46	1.22
SD	1.30	3.45	3.38	8.79	4.02
R^2	0.91	0.96	0.93	0.91	0.90
Number of Data	51	51	51	51	51
Total Set					
RMSE	0.38	0.44	0.69	0.59	0.76
SD	1.63	2.89	6.87	6.75	5.29
R^2	0.97	0.96	0.89	0.94	0.93
Number of Data	171	171	171	171	171

To provide a visual insight of permeability prediction by all methods, the plot of measured core data versus predicted data in the part of the dataset in which the used methods perform best is illustrated in Figure 10. This figure depicts predicted data (vertical axes), measured data (horizontal axes), and their regression plot. In this figure, permeability values of test data (red markers) and train data (blue markers) have been illustrated in graphical form. This can be deduced that the SVM model shows the highest match between measured and predicted data.



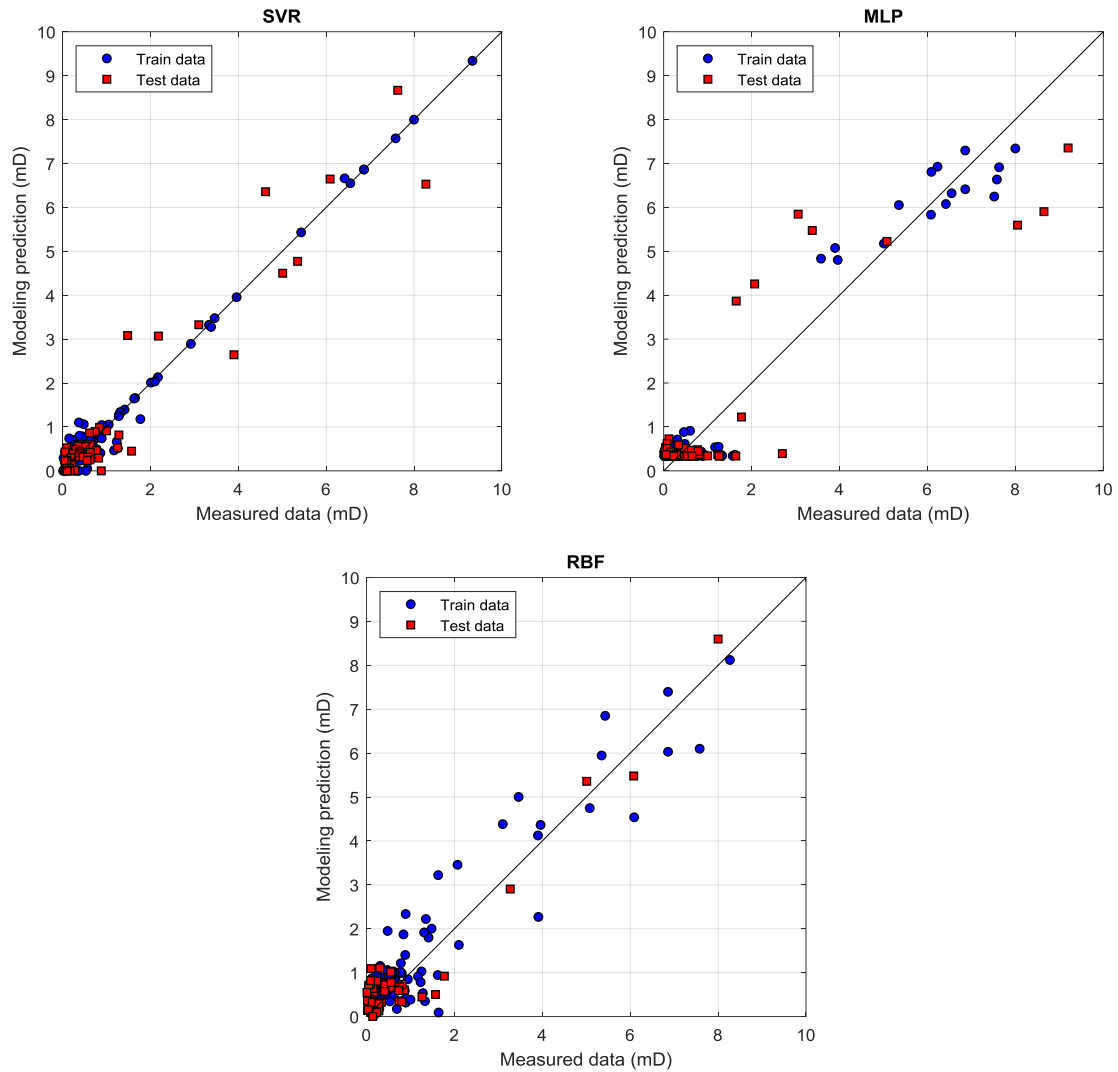
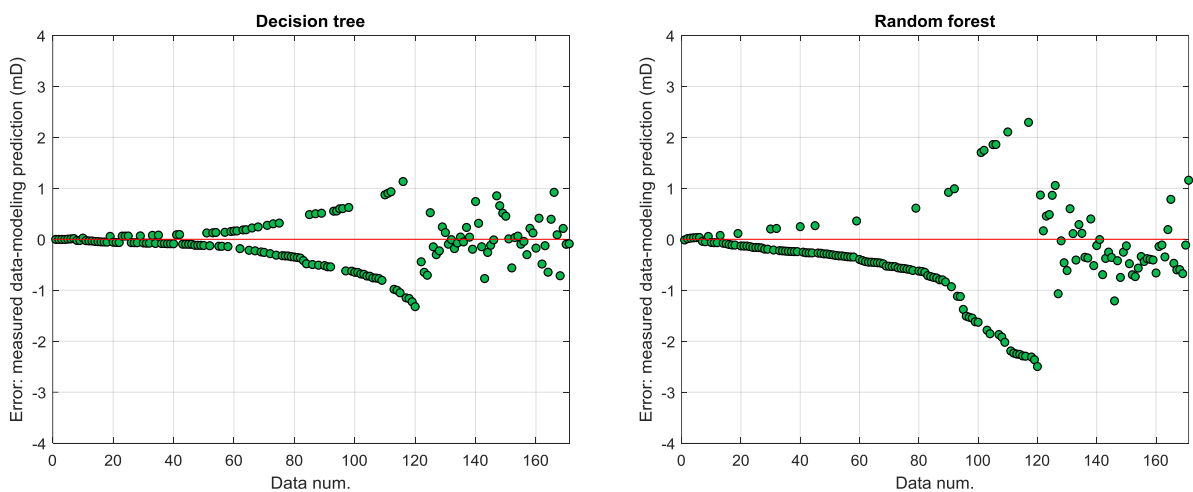


Figure 10. The cross plot of modeling prediction of permeability versus measured data for both train and test data

Moreover, Figure 11 illustrates the relative deviations of these models. In other words, these plots show the dispersion between data and outputs of constructed models. These variations can be distinguished in different ranges of data (horizontal axes). As illustrated in Figure 11, the applied SVM model shows the lowest data scattering and more concentrated data.



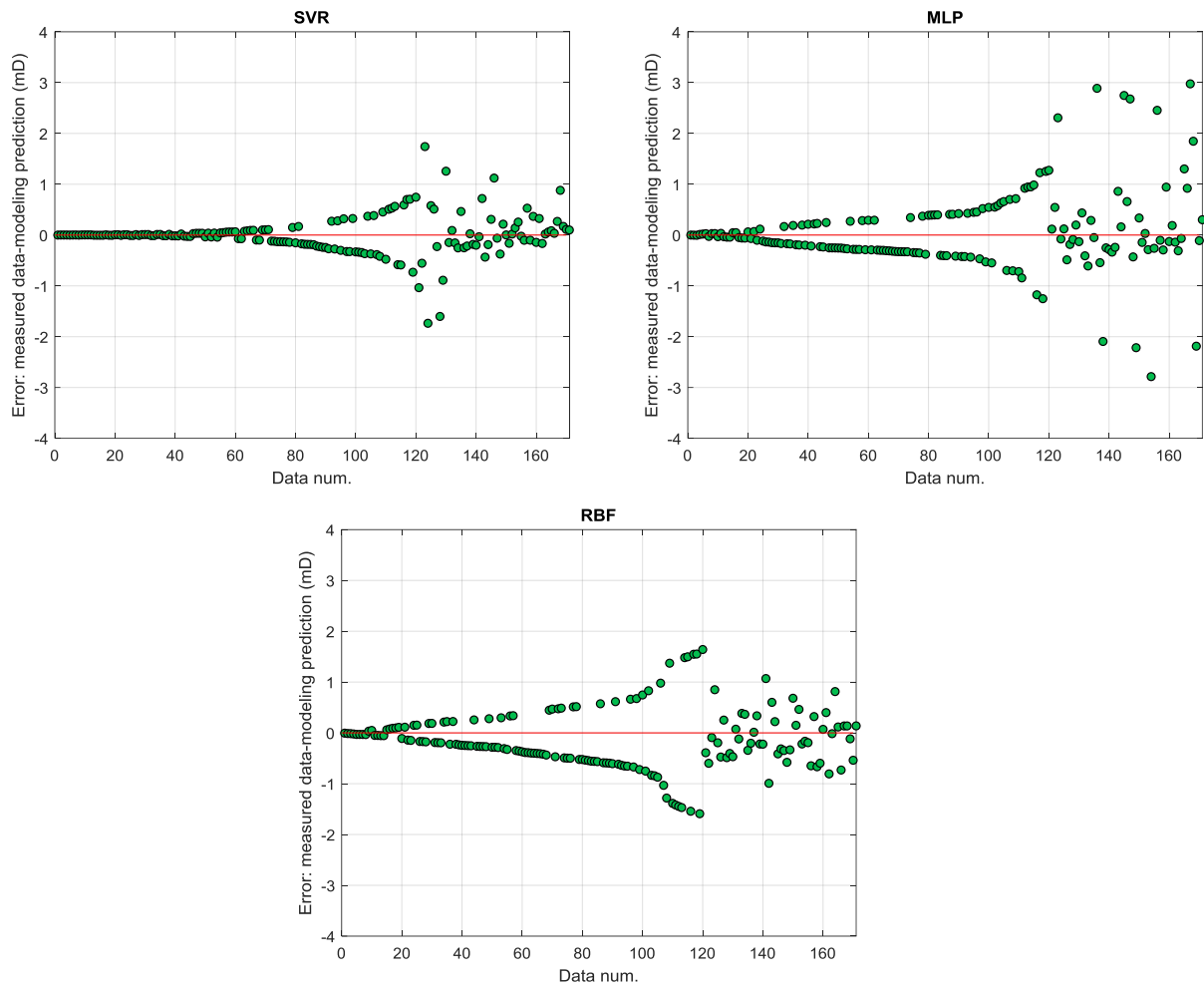


Figure 11. Relative deviations of DT, RF, SVR, MLP, and RBF models

For better comprehension, the performances of all methods have been depicted in Figure 12 to assess the differences visually. In this figure, the measured permeability values of core data are shown with black dots. Colored lines also offer the obtained permeability values by the other five methods. This figure represents the superiority of DT and SVM techniques for permeability prediction. As shown, this figure also demonstrates the weaker performance of the RF model to predict permeability values.

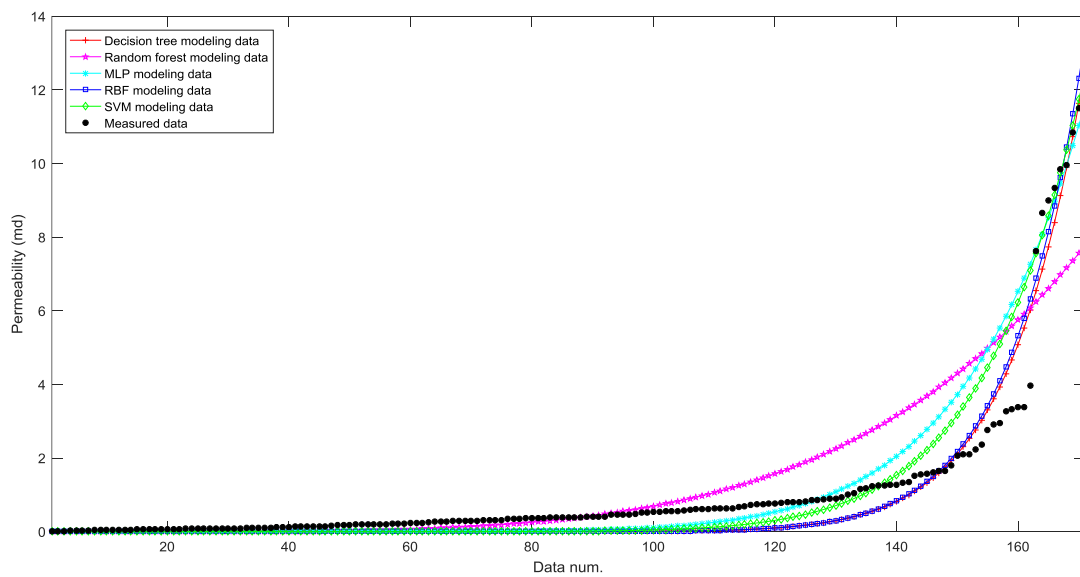


Figure 12. Comparison of different methods for prediction of permeability (black dots are measured data)

The use of experimental equations to estimate permeability has a long history [71-75]. These correlations relate the permeability values to pore characteristics such as porosity. The proposed empirical equations are not very precise, especially in carbonate rocks associated with the complicated texture. Also, these equations are merely applicable in the specified region. In other words, the generalization of these relations seems impossible [73]. In this section of this research, the achievement of this paper should be compared with previous experimental equations to assessing the preference of artificial intelligence methods. For this purpose, two proposed equations for permeability prediction in Ilam and Sarvak formations in this well has been reported as follow [76-78]:

$$\text{Ilam Formation} \quad K = 0/0113 \exp^{0/236\varphi} \quad (7)$$

$$\text{Sarvak Formation} \quad K = 0/1505 \exp^{0/2109\varphi} \quad (8)$$

Results of predicted permeability values based on these two aforementioned equations as well as the best-proposed models for permeability prediction in this study (DT and SVM), are illustrated in Figure 13. This figure provides a visual insight into comparison. As presented, these two graphs have been plotted against depth. Permeability values are shown in yellow circles, whereas predicted amounts of this parameter by equations and artificial intelligence approaches have been shown in colored lines. It must be emphasized that there is a gap in the range of 3200 m to 3235 m in the Sarvak Formation since core data has not been reported in this interval. Therefore, Sarvak Formation has been assessed in two distinct zones.

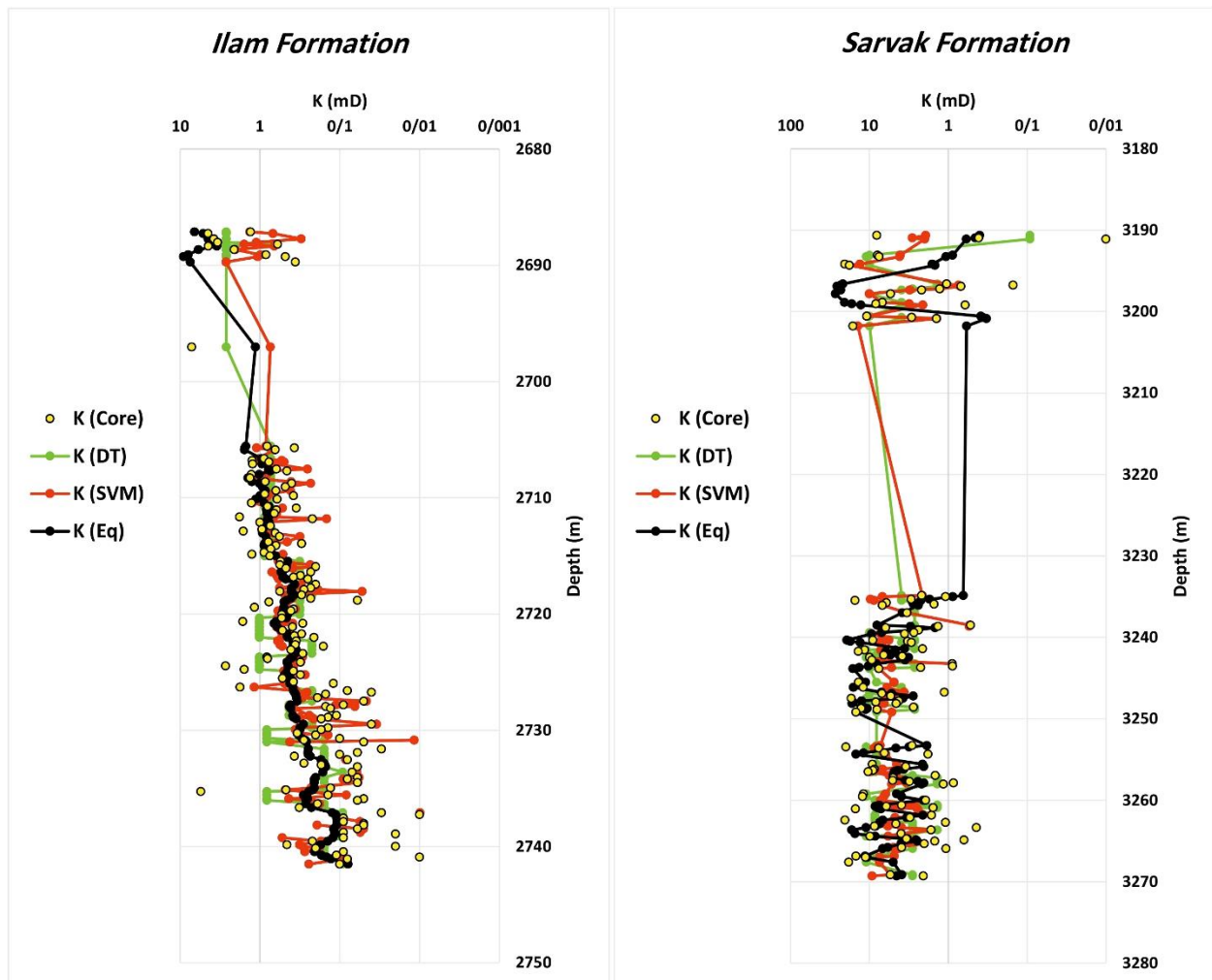


Figure 13. The variation of obtained permeability values by the empirical equation, SVM, and DT quantities with depth for Ilam and Sarvak formations

As depicted in Figure 13, experimental equations' performance is weaker than DT and SVM models in Ilam and Sarvak formations. These equations are incapable of predicting permeability values that are far from the average of the data, whereas DT and SVM models act reliably even in outlier data. These two artificial intelligence methods follow the trend of measured data appropriately. Moreover, statistical parameters of R^2 , RMSE and SD are calculated for predicted permeability by two equations of 7 and 8 in all over the Ilam and Sarvak formations. These values are listed in Table 4. For better comparison, the results of statistical parameters belong to DT and SVM methods are also listed in this paper.

As shown, statistical values vary dramatically in the equation case. R^2 exhibits an inappropriate and very low amount. RMSE and SD of equation case possess larger values than DT and SVM models too. These results indicate poor performance of empirical equations compared to artificial intelligence approaches.

Table 4. The calculated statistical parameters for the empirical equation as well as DT and SVM models

	Equation	SVM	DT
R^2	-0.75	0.97	0.96
RMSE	5.78	0.38	0.44
SD	11.26	1.63	2.89

5. Conclusion

Different types of data mining and neural network techniques were applied to predict the permeability values of the Ilam and Sarvak formations, which are two reservoirs in the southwestern part of Iran. For this target, MLP, RBF, and SVR neural networks, in addition to DT and RF techniques, were selected for an all-around comparison. Since other researchers investigated these methods individually, a comparison of several techniques can be considered the main novelty of the present paper. Well-logging data is used as input data. In the studied well, CGR, SGR, DT, LLD, LLS, MSFL, NPHI, RHOB, POTA, URAN, TH, and PEF logs are available. Random forest: selecting the variables with the most influence on permeability and reducing dimension was made by random forest. Depth, SGR, CGR, NPHI, and RHOB were chosen as the most influential parameters and were considered input data. 70% of the data was allocated for training and the rest for testing. The statistical parameters, such as R^2 , SD, and RMSE, were calculated for all approaches. Furthermore, the cross plots of measured and predicted data, in addition to the relative deviations of all models, were plotted. The graphical and statistical results of the constructed models show that all of them perform almost accurately. However, DT and SVM among these models are the best techniques for permeability prediction in the Ilam and Sarvak formations. The R^2 values of these two methods are 0.97 (SVM model) and 0.96 (DT model). Moreover, the SD and RMSE of these models have been calculated at 1.63 and 0.38 (for SVM) and 2.89 and 0.44 (for DT), respectively. The performances of two DT and SVM models in permeability prediction were compared with the previously proposed equation for this purpose. The results indicate empirical equations perform weakly in permeability prediction (with statistical parameters of $R^2 = -0.75$, $SD = 11.26$, and $RMSE = 5.78$), whereas machine learning methods act more accurately.

6. Declarations

6.1. Author Contributions

Conceptualization, M.T., Z.S. and M.S.; methodology, M.T., Z.S. and M.S.; software, M.T.; validation, M.T., Z.S. and M.S.; formal analysis, M.T.; investigation, M.T., Z.S.; resources, M.T., Z.S.; data curation, M.T., Z.S. and M.S.; writing—original draft preparation, Z.S., M.S.; writing—review and editing, Z.S., M.S.; visualization, M.T.; supervision, M.S.; project administration, M.S.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available in article.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Acknowledgements

The authors thank Ms. Farzaneh Talebkeikhah for her skilled technical assistance.

6.5. Institutional Review Board Statement

Not applicable.

6.6. Informed Consent Statement

Not applicable.

6.7. Declaration of Competing Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7. References

- [1] Aminian, K., Ameri, S., Oyerokun, A., & Thomas, B. (2003). Prediction of flow units and permeability using artificial neural networks. SPE Western Regional/AAPG Pacific Section Joint Meeting, 299–305. doi:10.2118/83586-ms.
- [2] Bagheripour, P. (2014). Committee neural network model for rock permeability prediction. *Journal of Applied Geophysics*, 104, 142–148. doi:10.1016/j.jappgeo.2014.03.001.
- [3] Lim, J. S. (2005). Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea. *Journal of Petroleum Science and Engineering*, 49(3–4), 182–192. doi:10.1016/j.petrol.2005.05.005.
- [4] Alnuaimi, M. M. (2018). Using Artificial Intelligence and Machine Learning to Develop Synthetic Well Logs Develop Synthetic Well Logs, West Virginia University, West Virginia, United States.
- [5] Ertekin, T., & Sun, Q. (2019). Use of Artificial Intelligence in Determining the Location of Infill Wells in Hydrocarbon Exploration and Production Activities. In *Advances in Petroleum Engineering and Petroleum Geochemistry* (pp. 3–5). Springer. doi:10.1007/978-3-030-01578-7_1.
- [6] Fajana, A. O., Ayuk, M. A., & Enikanselu, P. A. (2019). Application of multilayer perceptron neural network and seismic multiattribute transforms in reservoir characterization of Pennay field, Niger Delta. *Journal of Petroleum Exploration and Production Technology*, 9(1), 31–49. doi:10.1007/s13202-018-0485-9.
- [7] Ghaffarkhah, A., Afrand, M., Talebkeikhah, M., Sehat, A. A., Moraveji, M. K., Talebkeikhah, F., & Arjmand, M. (2020). On evaluation of thermophysical properties of transformer oil-based nanofluids: A comprehensive modeling and experimental study. *Journal of Molecular Liquids*, 300, 112249. doi:10.1016/j.molliq.2019.112249.
- [8] Khamis, M. A., & Fattah, K. A. (2019). Estimating oil–gas ratio for volatile oil and gas condensate reservoirs: artificial neural network, support vector machines and functional network approach. *Journal of Petroleum Exploration and Production Technology*, 9(1), 573–582. doi:10.1007/s13202-018-0501-0.
- [9] Moussa, T., Elkatatny, S., Mahmoud, M., & Abdulraheem, A. (2018). Development of New Permeability Formulation from Well Log Data Using Artificial Intelligence Approaches. *Journal of Energy Resources Technology, Transactions of the ASME*, 140(7). doi:10.1115/1.4039270.
- [10] Sabah, M., Talebkeikhah, M., Agin, F., Talebkeikhah, F., & Hasheminasab, E. (2019). Application of decision tree, artificial neural networks, and adaptive neuro-fuzzy inference system on predicting lost circulation: A case study from Marun oil field. *Journal of Petroleum Science and Engineering*, 177, 236–249. doi:10.1016/j.petrol.2019.02.045.
- [11] Salgado, W. L., Dam, R. S. de F., Teixeira, T. P., Conti, C. C., & Salgado, C. M. (2020). Application of artificial intelligence in scale thickness prediction on offshore petroleum using a gamma-ray densitometer. *Radiation Physics and Chemistry*, 168, 108549. doi:10.1016/j.radphyschem.2019.108549.
- [12] Sadeghtabaghi, Z., Talebkeikhah, M., & Rabbani, A. R. (2020). Prediction of vitrinite reflectance values using machine learning techniques: a new approach. *Journal of Petroleum Exploration and Production Technology*, 11(2), 651–671. doi:10.1007/s13202-020-01043-8.
- [13] Bhatt, A., & Helle, H. B. (2002). Committee neural networks for porosity and permeability prediction from well logs. *Geophysical Prospecting*, 50(6), 645–660. doi:10.1046/j.1365-2478.2002.00346.x.
- [14] Elkatatny, S., Mahmoud, M., Tariq, Z., & Abdulraheem, A. (2018). New insights into the prediction of heterogeneous carbonate reservoir permeability from well logs using artificial intelligence network. *Neural Computing and Applications*, 30(9), 2673–2683. doi:10.1007/s00521-017-2850-x.
- [15] Helle, H. B., Bhatt, A., & Ursin, B. (2001). Porosity and permeability prediction from wireline logs using artificial neural networks: A North Sea case study. *Geophysical Prospecting*, 49(4), 431–444. doi:10.1046/j.1365-2478.2001.00271.x.
- [16] Zhong, Z., Carr, T. R., Wu, X., & Wang, G. (2019). Application of a convolutional neural network in permeability prediction: A case study in the Jacksonburg-Stringtown oil field, West Virginia, USA. *Geophysics*, 84(6), B363–B373. doi:10.1190/geo2018-0588.1.
- [17] Chen, C. H., & Lin, Z. S. (2006). A committee machine with empirical formulas for permeability prediction. *Computers and Geosciences*, 32(4), 485–496. doi:10.1016/j.cageo.2005.08.003.
- [18] Karimpouli, S., Fathianpour, N., & Roohi, J. (2010). A new approach to improve neural networks' algorithm in permeability prediction of petroleum reservoirs using supervised committee machine neural network (SCMNN). *Journal of Petroleum Science and Engineering*, 73(3–4), 227–232. doi:10.1016/j.petrol.2010.07.003.
- [19] Gholami, R., Moradzadeh, A., Maleki, S., Amiri, S., & Hanachi, J. (2014). Applications of artificial intelligence methods in prediction of permeability in hydrocarbon reservoirs. *Journal of Petroleum Science and Engineering*, 122, 643–656. doi:10.1016/j.petrol.2014.09.007.

- [20] Shokooh Saljooghi, B., & Hezarkhani, A. (2015). A new approach to improve permeability prediction of petroleum reservoirs using neural network adaptive wavelet (wavenet). *Journal of Petroleum Science and Engineering*, 133, 851–861. doi:10.1016/j.petrol.2015.04.002.
- [21] Jamshidian, M., Hadian, M., Zadeh, M. M., Kazempoor, Z., Bazargan, P., & Salehi, H. (2015). Prediction of free flowing porosity and permeability based on conventional well logging data using artificial neural networks optimized by Imperialist competitive algorithm - A case study in the South Pars gas field. *Journal of Natural Gas Science and Engineering*, 24, 89–98. doi:10.1016/j.jngse.2015.02.026.
- [22] Handhal, A. M. (2016). Prediction of reservoir permeability from porosity measurements for the upper sandstone member of Zubair Formation in Super-Giant South Rumila oil field, southern Iraq, using M5P decision tree and adaptive neuro-fuzzy inference system (ANFIS): a comparative study. *Modeling Earth Systems and Environment*, 2(3), 111. doi:10.1007/s40808-016-0179-6.
- [23] Rafik, B., & Kamel, B. (2017). 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 Journal of Petroleum*, 26(3), 763–778. doi:10.1016/j.ejpe.2016.10.013.
- [24] Al-Amri, M., Mahmoud, M., Elkatatny, S., Al-Yousef, H., & Al-Ghamdi, T. (2017). Integrated petrophysical and reservoir characterization workflow to enhance permeability and water saturation prediction. *Journal of African Earth Sciences*, 131, 105–116. doi:10.1016/j.jafrearsci.2017.04.014.
- [25] Zhao, T., Zhao, H., Ning, Z., Li, X., & Wang, Q. (2018). Permeability prediction of numerical reconstructed multiscale tight porous media using the representative elementary volume scale lattice Boltzmann method. *International Journal of Heat and Mass Transfer*, 118, 368–377. doi:10.1016/j.ijheatmasstransfer.2017.11.004.
- [26] Wood, D. A. (2020). Predicting porosity, permeability and water saturation applying an optimized nearest-neighbour, machine-learning and data-mining network of well-log data. *Journal of Petroleum Science and Engineering*, 184, 106587. doi:10.1016/j.petrol.2019.106587.
- [27] Jooybari, A., & Rezaie, P. (2017). Petrophysical Evaluation of the Sarvak Formation Based on Well Logs in Dezful Embayment, Zagros Fold Zone, South West of Iran. *Engineering, Technology & Applied Science Research*, 7(1), 1358–1362. doi:10.48084/etasr.982.
- [28] Nabikhani, N., Mahboubi, A., Kadhodaie, A., & Yosefpour, M. R. (2012). The Evaluation of Reservoir Quality of Sarvak Formation in One of Oil Fields of the Persian Gulf. *Journal of Petroleum Science and Technology*, 2(1), 2012.
- [29] Poursamad, R., & Tabatabaei, H. (2017). Reservoir Quality Evaluation of Sarvak Formation in Gachsaran Oil Field, SW of Iran (Issue June). SW of Iran.
- [30] Beydoun, Z. R. (1988). The Middle East: regional geology and petroleum resources. In *The Middle East: regional geology and petroleum resources*. doi:10.2307/635079.
- [31] Alizadeh, B., Saadati, H., Rashidi, M., & Kobraei, M. (2016). Geochemical investigation of oils from Cretaceous to Eocene sedimentary sequences of the Abadan Plain, Southwest Iran. *Marine and Petroleum Geology*, 73, 609–619. doi:10.1016/j.marpetgeo.2015.11.002.
- [32] Saadatinejad, M. R., & Sarkarinejad, K. (2011). Application of the spectral decomposition technique for characterizing reservoir extensional system in the Abadan Plain, southwestern Iran. *Marine and Petroleum Geology*, 28(6), 1205–1217. doi:10.1016/j.marpetgeo.2011.02.002.
- [33] Assadi, A., Honarmand, J., Moallemi, S. A., & Abdollahie-Fard, I. (2016). Depositional environments and sequence stratigraphy of the Sarvak Formation in an oil field in the Abadan Plain, SW Iran. *Facies*, 62(4), 26. doi:10.1007/s10347-016-0477-5.
- [34] Saadatmand, M. R., Moradi, A., & Hashemi, H. (2013). Passive seismic survey on the Darquain oil field. *Quarterly Journal of Tethys*, 1(3), 215–224.
- [35] Rajabi, M., Sherkati, S., Bohloli, B., & Tingay, M. (2010). Subsurface fracture analysis and determination of in-situ stress direction using FMI logs: An example from the Santonian carbonates (Ilam Formation) in the Abadan Plain, Iran. *Tectonophysics*, 492(1–4), 192–200. doi:10.1016/j.tecto.2010.06.014.
- [36] G. A. James (2), J. G. Wynd (3). (1965). Stratigraphic Nomenclature of Iranian Oil Consortium Agreement Area. *AAPG Bulletin*, 49(12), 2182–2245. doi:10.1306/a663388a-16c0-11d7-8645000102c1865d.
- [37] Assadi, A., Honarmand, J., Moallemi, S. A., & Abdollahie-Fard, I. (2018). An integrated approach for identification and characterization of palaeo-exposure surfaces in the upper Sarvak Formation of Abadan Plain, SW Iran. *Journal of African Earth Sciences*, 145, 32–48. doi:10.1016/j.jafrearsci.2018.05.002.
- [38] Rezaie, P., Jooybari, A., Pour, M. M., & Gorbani, M. (2016). Factor Controlling Reservoir Properties and Flow Unit Determination in the Ilam Formation of Dezful Embayment at Zagros Fold-Thrust Belt, Southwest of Iran. *Open Journal of Geology*, 06(07), 660–671. doi:10.4236/ojg.2016.67051.

- [39] Esrafil-Dizaji, B., Rahimpour-Bonab, H., Mehrabi, H., Afshin, S., Kiani Harchegani, F., & Shahverdi, N. (2015). Characterization of rudist-dominated units as potential reservoirs in the middle Cretaceous Sarvak Formation, SW Iran. *Facies*, 61(3), 14. doi:10.1007/s10347-015-0442-8.
- [40] Kosari, E., Kadjhodaie, A., Bahroudi, A., Chehrizi, A., & Talebian, M. (2017). An integrated approach to study the impact of fractures distribution on the Ilam-Sarak carbonate reservoirs: A case study from the Strait of Hormuz, the Persian Gulf. *Journal of Petroleum Science and Engineering*, 152, 104–115. doi:10.1016/j.petrol.2017.03.001.
- [41] Nasser, A., Mohammadzadeh, M. J., & Hashemtabatabaee, S. (2016). Evaluating Bangestan reservoirs and targeting productive zones in Dezful embayment of Iran. *Journal of Geophysics and Engineering*, 13(6), 994–1001. doi:10.1088/1742-2132/13/6/994.
- [42] Deosarkar, M. P., & Sathe, V. S. (2012). Predicting effective viscosity of magnetite ore slurries by using artificial neural network. *Powder Technology*, 219, 264–270. doi:10.1016/j.powtec.2011.12.058.
- [43] Azimi Djivejin, Z., Ghaffarkhah, A., Vafaie Sefti, M., & Moraveji, M. K. (2019). Synthesis, structure and mechanical properties of nanocomposites based on exfoliated nano magnesium silicate crystal and poly(acrylamide). *Journal of Dispersion Science and Technology*, 40(2), 276–286. doi:10.1080/01932691.2018.1467777.
- [44] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer Texts in Statistics. doi:10.1007/978-1-4614-7138-7.
- [45] Paławski, A., Szłęk, J., Lau, R., Jachowicz, R., & Mendyk, A. (2015). Empirical modeling of the fine particle fraction for carrier-based pulmonary delivery formulations. *International Journal of Nanomedicine*, 10, 801–810. doi:10.2147/IJN.S75758.
- [46] Maučec, M., Singh, A. P., Bhattacharya, S., Yarus, J. M., Fulton, D. D., & Orth, J. M. (2015). Multivariate analysis and data mining of well-stimulation data by use of classification-and-regression tree with enhanced interpretation and prediction capabilities. *SPE Economics and Management*, 7(2), 60–71. doi:10.2118/166472-PA.
- [47] Singh, A. (2017). Application of data mining for quick root-cause identification and automated production diagnostic of gas wells with plunger lift. *SPE Production and Operations*, 32(3), 279–293. doi:10.2118/175564-pa.
- [48] Singh, A. (2015). Root-cause identification and production diagnostic for gas wells with plunger lift. In *Society of Petroleum Engineers - SPE Reservoir Characterisation and Simulation Conference and Exhibition, RCSC*, 1042–1064. doi:10.2118/175564-ms.
- [49] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and Regression Trees*. Routledge. doi:10.1201/9781315139470.
- [50] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32. doi:10.1023/A:1010933404324.
- [51] Eskandarian, S., Bahrami, P., & Kazemi, P. (2017). A comprehensive data mining approach to estimate the rate of penetration: Application of neural network, rule based models and feature ranking. *Journal of Petroleum Science and Engineering*, 156, 605–615. doi:10.1016/j.petrol.2017.06.039.
- [52] Hegde, C., Wallace, S., & Gray, K. (2015). Using trees, bagging, and random forests to predict rate of penetration during drilling. *Society of Petroleum Engineers - SPE Middle East Intelligent Oil and Gas Conference and Exhibition*. doi:10.2118/176792-ms.
- [53] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees* (Chapman & Hall, eds.). Monterey, CA, EE. UU: Wadsworth International Group.
- [54] Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22.
- [55] Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, 147, 77–89. doi:10.1016/j.enbuild.2017.04.038.
- [56] Okpo, E. E., Dosunmu, A., & Odagme, B. S. (2016). Artificial neural network model for predicting wellbore instability. *Society of Petroleum Engineers - SPE Nigeria Annual International Conference and Exhibition*. doi:10.2118/184371-ms.
- [57] Al-Naser, M., Elshafei, M., & Al-Sarkhi, A. (2016). Artificial neural network application for multiphase flow patterns detection: A new approach. *Journal of Petroleum Science and Engineering*, 145, 548–564. doi:10.1016/j.petrol.2016.06.029.
- [58] Khoshjavan, S., Heidary, M., & Rezai, B. (2010). Estimation of coal swelling index based on chemical properties of coal using artificial neural networks. *Iranian Journal of Materials Science and Engineering*, 7(3), 1–11.
- [59] Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14–15), 2627–2636. doi:10.1016/S1352-2310(97)00447-0.
- [60] Moghadassi, A., Parvizian, F., & Hosseini, S. M. (2009). A new approach based on artificial neural networks for prediction of high pressure vapor-liquid equilibrium. *Australian Journal of Basic and Applied Sciences*, 3(3), 1851–1862.
- [61] Hemmati-Sarapardeh, A., Ghazanfari, M. H., Ayatollahi, S., & Masihi, M. (2016). Accurate determination of the CO₂-crude oil minimum miscibility pressure of pure and impure CO₂ streams: A robust modelling approach. *Canadian Journal of Chemical Engineering*, 94(2), 253–261. doi:10.1002/cjce.22387.

- [62] Lashkarbolooki, M., Hezave, A. Z., & Ayatollahi, S. (2012). Artificial neural network as an applicable tool to predict the binary heat capacity of mixtures containing ionic liquids. *Fluid Phase Equilibria*, 324, 102–107. doi:10.1016/j.fluid.2012.03.015.
- [63] Yilmaz, I., & Kaynar, O. (2011). Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. *Expert Systems with Applications*, 38(5), 5958–5966. doi:10.1016/j.eswa.2010.11.027.
- [64] Hemmat Esfe, M., Saedodin, S., Sina, N., Afrand, M., & Rostami, S. (2015). Designing an artificial neural network to predict thermal conductivity and dynamic viscosity of ferromagnetic nanofluid. *International Communications in Heat and Mass Transfer*, 68, 50–57. doi:10.1016/j.icheatmasstransfer.2015.06.013.
- [65] Huang, S. C., & Do, B. H. (2014). Radial basis function based neural network for motion detection in dynamic scenes. *IEEE Transactions on Cybernetics*, 44(1), 114–125. doi:10.1109/TCYB.2013.2248057.
- [66] Venkatesan, P., & Anitha, S. (2006). Application of a radial basis function neural network for diagnosis of diabetes mellitus. *Current Science*, 91(9), 1195–1199.
- [67] Hemmati-Sarapardeh, A., Varamesh, A., Husein, M. M., & Karan, K. (2018). On the evaluation of the viscosity of nanofluid systems: Modeling and data assessment. *Renewable and Sustainable Energy Reviews*, 81, 313–329. doi:10.1016/j.rser.2017.07.049.
- [68] Panda, S. S., Chakraborty, D., & Pal, S. K. (2008). Flank wear prediction in drilling using back propagation neural network and radial basis function network. *Applied Soft Computing Journal*, 8(2), 858–871. doi:10.1016/j.asoc.2007.07.003.
- [69] Nasiri, A., Ghaffarkhah, A., Keshavarz Moraveji, M., Gharbanian, A., & Valizadeh, M. (2017). Experimental and field test analysis of different loss control materials for combating lost circulation in bentonite mud. *Journal of Natural Gas Science and Engineering*, 44, 1–8. doi:10.1016/j.jngse.2017.04.004.
- [70] Hemmati-Sarapardeh, A., Khishvand, M., Naseri, A., & Mohammadi, A. H. (2013). Toward reservoir oil viscosity correlation. *Chemical Engineering Science*, 90, 53–68. doi:10.1016/j.ces.2012.12.009.
- [71] Altunbay, M., Georgi, D., & Takezaki, H. M. (1997). Permeability prediction for carbonates: still a challenge? In *Proceedings of the Middle East Oil Show*, Vol. 1, 609–620. doi:10.2523/37753-ms.
- [72] Amaefule, J. O., Altunbay, M., Tiab, D., Kersey, D. G., & Keelan, D. K. (1993). Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/ wells. In *Proceedings - SPE Annual Technical Conference and Exhibition: Vol. Omega*, 205–220. doi:10.2523/26436-ms.
- [73] Babadagli, T., & Al-Salmi, S. (2002). Improvement of Permeability Prediction for Carbonate Reservoirs Using Well Log Data. In *SPE - Asia Pacific Oil and Gas Conference*, 555–571. doi:10.2118/77889-ms.
- [74] Carman, P. G. (1997). Fluid flow through granular beds. *Chemical Engineering Research and Design*, 75(1 SUPPL.), 32–48. doi:10.1016/s0263-8762(97)80003-2.
- [75] Ohen, H. A., Ajufo, A., & Curby, F. M. (1995). Hydraulic (flow) unit based model for the determination of petrophysical properties from NMR relaxation measurements. In *Proceedings - SPE Annual Technical Conference and Exhibition: Vol. Omega*, 983–996. doi:10.2523/30626-ms.
- [76] Shabani, M., Sadeghtabaghi, Z., & Khoshshar, Z. (2020). Petrophysical Evaluation of Bangestan Group Formations in an Iranian Oil Field. *Journal of Oil, Gas and Petrochemical Technology*, 7(1), 30–42.
- [77] Hemmat Esfe, M., Wongwises, S., Naderi, A., Asadi, A., Safaei, M. R., Rostamian, H., Dahari, M., & Karimipour, A. (2015). Thermal conductivity of Cu/TiO₂-water/EG hybrid nanofluid: Experimental data and modeling using artificial neural network and correlation. *International Communications in Heat and Mass Transfer*, 66, 100–104. doi:10.1016/j.icheatmasstransfer.2015.05.014.
- [78] Vapnik, V. (2013). *The nature of statistical learning theory*: Springer Science & Business Media, Berlin, Germany.