ORIGINAL PAPER-EXPLORATION GEOLOGY

Prediction of vitrinite reflectance values using machine learning techniques: a new approach

Zahra Sadeghtabaghi[1] · Mohsen Talebkeikhah[1] · Ahmad Reza Rabbani[1]

Abstract

Vitrinite reflectance (VR) is considered the most used maturity indicator of source rocks. Although vitrinite reflectance is an acceptable parameter for maturity and is widely used, it is sometimes difficult to measure. Furthermore, Rock-Eval pyrolysis is a current technique for geochemical investigations and evaluating source rock by their quality and quantity of organic matter, which provide low cost, quick, and valid information. Predicting vitrinite reflectance by using a quick and straightforward method like Rock-Eval pyrolysis results in determining accurate and reliable values of VR with consuming low cost and time. Previous studies used empirical equations for vitrinite reflectance prediction by the T_{max} data, which was accompanied by poor results. Therefore, finding a way for precise vitrinite reflectance prediction by Rock-Eval data seems useful. For this aim, vitrinite reflectance values are predicted by 15 distinct machine learning models of the decision tree, random forest, support vector machine, group method of data handling, radial basis function, multilayer perceptron, adaptive neuro-fuzzy inference system, and multilayer perceptron and adaptive neuro-fuzzy inference system, which are coupled with evolutionary optimization methods such as grasshopper optimization algorithm, bat algorithm, particle swarm optimization, and genetic algorithm, with four inputs of Rock-Eval pyrolysis parameters of T_{max}, S_1/TOC, HI, and depth for the first time. Statistical evaluations indicate that the decision tree is the most precise model for VR prediction, which can estimate vitrinite reflectance precisely. The comparison between the decision tree and previous proposed empirical equations indicates that the machine learning method performs much more accurately.

Keywords Vitrinite reflectance · Rock-eval pyrolysis · Maturity · Depth · Decision tree

Introduction

Rock-Eval pyrolysis is considered one of the most powerful and influential geochemical techniques which can provide valuable information about organic matter sediments promptly (Espitalié et al. 1977; Behar et al. 2001). This technique is also used extensively in petroleum source rock evaluation through the determination of organic matter types, generation potential, and level of maturity (Dembicki 2016). S_1 (formerly generated hydrocarbon), S_2 (remain potential of hydrocarbon generation), TOC (total organic carbon), T_{max} (temperature of S_2 maximum), HI (hydrogen index), and OI (oxygen index) are some parameters which are obtained by Rock-Eval pyrolysis. Despite all Rock-Eval advantages, this method is associated with some obstacles for maturity interpretations as well as the determination of organic matter type (Katz 1983). T_{max} is a parameter which varies with depth and is used for thermal maturity determination (Espitalié et al. 1977). Changing of Rock-Eval apparatus condition, the matrix of analyzed sample, and heavy components of bitumen are factors which influence on T_{max} determination. As a consequence, this parameter can sometimes result in inverted and fallacious conclusions toward the actual level of maturity (Espitalié 1986). Furthermore, in situations in which the S_2 peak has more than one maximum, or suffers weak intensity, T_{max} determination will be complicated (Peters 1986). It should also be noted that maturity determination through the T_{max} parameter is changed by some other factors such as bitumen or kerogen in adjacent units (Snowdon 1995). Therefore, T_{max} can't be considered as a dependable indicator for thermal maturity interpretation. In the cases that the T_{max} is affected by the facies effects,

✉ Zahra Sadeghtabaghi
  zahrasadeghtabaghi@aut.ac.ir

[1] Department of Petroleum Engineering, Amir Kabir University of Technology, Tehran, Iran

vitrinite reflectance can provide more reliable results (Peters and Cassa 1994).

In addition to the Rock-Eval parameters, time alteration index (TAI), conodont alteration index (CAI), and vitrinite reflectance (VR) are considered as other maturity indicators. TAI is defined based on color variations of kerogen. The color range of spore and pollen from yellow to black (immature to mature) is then converted to a numerical scale. Since spore and pollen belong to terrigenous plants that are absent before the middle Paleozoic, this parameter has some problems (Staplin 1961). Also, CAI is defined similarly. The color of conodonts is studied by binocular microscopes. These types of fossils are found in carbonate rocks so this index is improper for maturity assessment of shale rocks. Moreover, CAI cannot be used in cases earlier than Jurassic (Epstein et al. 1976). Zooclast and solid bitumen reflectivities are 2 other known petrographic methods for this aim. It should be noted that these methods are not free of problems. The identification of the correct Zooclast is always difficult. Also, different sources cause this problem in solid bitumen assessment (Bertrand 1990; Cole 1994; Petersen et al. 2013; Curiale 1986).

Vitrinite particle as a diagenetic product of higher plant remnant is one of the most significant macerals of petroleum source rock that can be found dispersed in other clastic sedimentary rocks and majorly in coals (Suggate 1959; Taylor et al. 1998). The light reflection phenomenon from a polished vitrinite surface is known as vitrinite reflectance (Mukhopadhyay 1994). Vitrinite reflectance (Ro%) is sensitive to temperature variations. In 1982, for the first time, it was discovered that vitrinite reflectance increases with time and temperature (Teichmuller and Teichmuller 1982). Thus, vitrinite reflectance varies based on maturity and is assumed to be the traditional and most robust diagnostic tool for maturity investigation, which can be applicable in a wide range of maturity levels (Mählmann and Le Bayon 2016). Data of this parameter can be applied for source rock evaluation or coal rank assessment (Kadkhodaie and Rezaee 2017; Jiang et al. 2019).

Vitrinite reflectance (VR) is the most applied factor for maturity determination (Cheshire et al. 2017; Peters et al. 2018). In addition to geological evaluation of petroleum source rocks as well as the coal rank assessment, VR possesses a significant role in basin modeling. Vitrinite reflectance data are the most known calibration parameters in the modeling procedure. Kinetic maturation of source rocks modeling also requires maturation data (Mukhopadhyay 1994). Vitrinite reflectance data are commonly used in three forms of empirical, single reaction kinetic, and parallel reactions kinetic (Barker and Elders 1981; Waples 1984; Welte and Yalcin 1988; Sweeney and Burnham 1990). The VR values of hydrocarbon generation in different types of kerogen can be helpful for source rock potential evaluation.

For instance, type II-S kerogen starts generating earlier than other types (Dembicki 2009). Moreover, the method of EASY %Ro as the most famous and reputable concept of maturation in basin modeling considers a defined range for activation energy distribution (Sweeney and Burnham 1990) and tectonic history determination (Middleton 1982) and seeking hidden pluton in sedimentary rocks (Chen et al. 2017).

According to the aforementioned issue, finding a way to relate T_{max} to vitrinite reflectance can cover deficits of imperfect maturity interpretations based on the T_{max} parameter and results in more precise outcomes. In other words, there are several associated problems with measuring vitrinite reflectance such as time-consuming, lack of vitrinite particles in some cases, and the anisotropy of vitrinite (Wust et al. 2013; Dembicki 2016). Also, the vitrinite maceral belongs to after Devonian plants, and samples that are older than this time don't have any vitrinite. In addition to problems with the age of source rock samples, low amounts of incoming plants result in low quantities of vitrinite values (Peters and Cassa 1994), and oxidation of this maceral also results in high values (Liu et al. 2020). In other words, two maturity indicators of VR and T_{max} have some problems, but the shortcoming of the first one is in measuring and the second one in the final result. Therefore, finding a way of predicting vitrinite reflectance through the simple procedure of Rock-Eval pyrolysis can be considered as a novel and efficient method that can dispel the problem of VR measuring by using several Rock-Eval parameters. Because of challenges in VR determination, finding a way for maturity estimation had been an interesting subject for researchers. Some of them have studied the relationship between T_{max} and VR parameters and proposed equations that can calculate vitrinite reflectance based on T_{max} data (Galimov and Rabbani 2001; Jarvie et al. 2001; Wust et al. 2013; Peters 1986; Jarvie 2012). It must be emphasized that all these equations suffer from some deficits, such as a low coefficient of determination or incorrect predictions in some cases. Also, these types of VR calculations are simple correlations just on the basis of one parameter (T_{max}), and this issue increases uncertainties. Some efforts have been made to determine vitrinite reflectance values by spectroscopic techniques which are also based on equations (Cheshmeh Sefidi and Ajorkaran 2019; Wilkins et al. 2015, 2018; Kibria et al. 2020). Moreover, Lupoi et al. (2019) and Hou et al. (2020) used the partial least squares model by Raman spectra of shale samples and modified the EASY% Ro model for vitrinite reflectance prediction, respectively.

The main objective of this paper is the precise prediction of vitrinite reflectance based on Rock-Eval parameters, not just T_{max} data or equation-based methods. Contrary to limited previous researches that have merely used the T_{max} parameter for vitrinite reflectance predictions or some simple

equations, this paper, for the first time, uses machine learning methods by using depth, T_{max}, S_1/TOC, and HI. Also, using machine learning for determining vitrinite reflectance values has been unprecedented. Previous researches for predicting geochemical parameters using machine learning methods were done to predict rock eval data such as TOC (Khoshnoodkia et al. 2011; Shalaby et al. 2019; Ge et al. 2015), $S1$, and $S2$ (Johnson et al. 2018; Wang et al. 2019). None of the earlier studies were about vitrinite reflectance prediction. Also, all machine learning methods in this article are assessed, which is unprecedented.

Artificial intelligence methods can learn the pattern of data and use multiple inputs, but previous attempts were merely based on one input without any learning (just simple regression). Also, this paper applies not only one way but also 15 distinct methods for this purpose. Moreover, the results of these methods are compared to finding the most proper model. The final result of the best machine learning method is then compared with the previously presented equations. The depth, T_{max}, S_1/TOC, and HI parameters are applied as inputs of constructed models by 15 machine learning methods of decision tree (DT), random forest (RF), support vector machine (SVM), group method of data handling (GMDH), radial basis function (RBF), multilayer perceptron (MLP), multilayer perceptron couple with grasshopper optimization algorithm (MLP+GOA), bat algorithm (MLP+BAT), particle swarm optimization algorithm (MLP+PSO), and genetic algorithm (MLP+GA) as well as adaptive neuro-fuzzy inference system (ANFIS), ANFIS + GOA, ANFIS + BAT, ANFIS + PSO and ANFIS + GA for VR determination. Then, the accuracy of all methods was compared through four statistical parameters of average absolute relative deviation (AARD), coefficient of determination (R2), root mean square error (RMSE),

and standard deviation (SD). Eventually, the most accurate model was selected. The final results of this paper were then compared with the results of previous efforts. A schematic form of the methodology of this paper is shown in Fig. 1.

Materials and methods

Data gathering and preparation

As mentioned in this paper, the required data for predicting VR are depth, T_{max}, S_1/TOC, and HI. To that end, 54 data sets of these aforementioned data are extracted from a broad area of the Persian Gulf as can be seen in Fig. 2. The Persian Gulf, as the most significant foreland basin
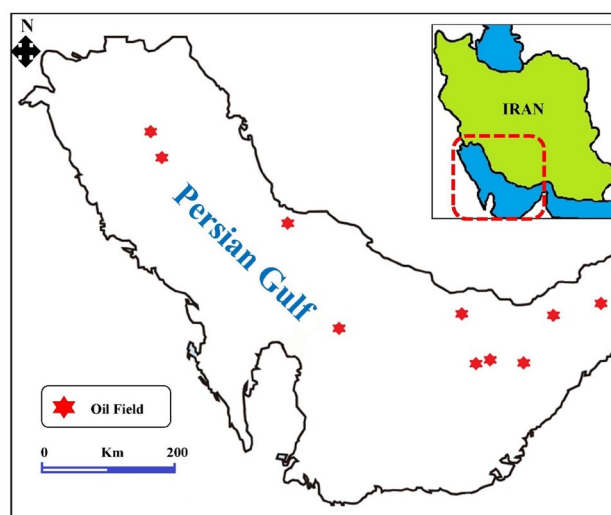
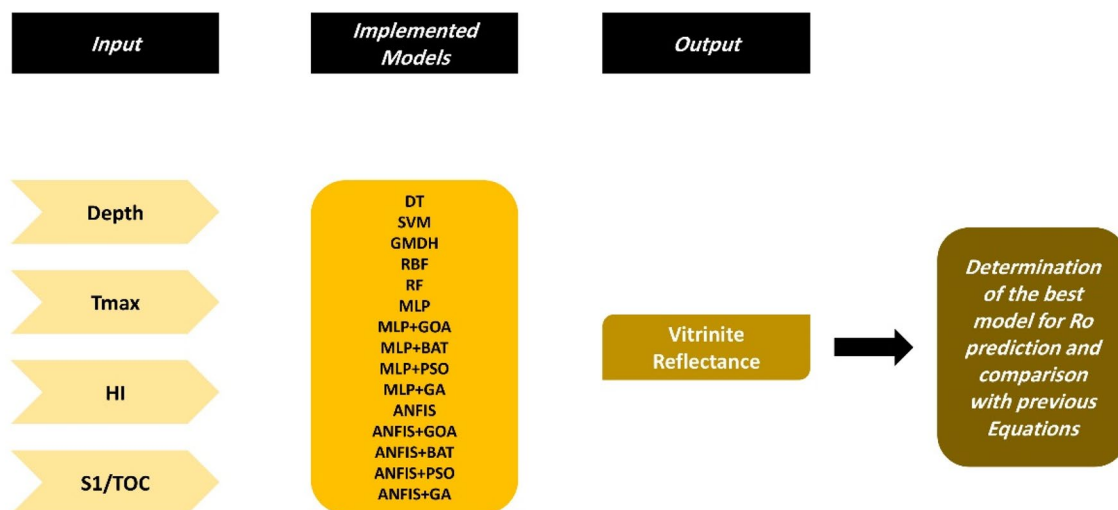**Fig. 2** Location map of studied wells in the Persian Gulf

Fig. 1 Schematic of applied methodology in this paper

of hydrocarbon resources in the world, is located at the confluence of Arabian and Eurasian plates. This contains more than two-thirds of the proven oil reserves of the world (Rabbani 2007; Haghi et al. 2013). Pabdeh (Paleocene), Gurpi (Campanian–Paleocene), Ahmadi member of Sarvak (Cenomanian), and Kazhdumi (Albian) are considered as the most probable source rocks of the Middle Cretaceous–Early Miocene petroleum system of the studied area (Mashhadi et al. 2015). These formations are illustrated in Fig. 3 in the form of a simple stratigraphic column. Pabdeh Formation consists mainly of shale, argillaceous limestone, and marls, which overlies Gurpi Formation. Gurpi is another probable source rock of this area that can also act as cap rock of lower reserves and includes marl and marly limestone. Ahmadi Member contains shaly facies, especially in northern parts of the Persian Gulf, and can be considered as one of the probable intervals for hydrocarbon generation. Eventually, Kazhdumi, as the most crucial source layer, generally consists of calcareous shale as well as dark bituminous limestone (Ghazban 2009; Homke et al. 2009; Soleimani et al. 2013).

After selecting proper and complete data, 54 Rock-Eval pyrolysis data (depth, T_{max}, S_1/TOC, and HI) belong to ten oil fields in the Iranian sector of the Persian Gulf (Fig. 2) were extracted (Mashhadi et al. 2015). Minimum, maximum, and average values of selected data are summarized in Table 1. Twenty-one samples belong to Pabdeh Formation, whereas contributions of Gurpi Formation, Ahmadi member, and Kazhdumi Formation are 6, 4, and 23 samples, respectively. In this paper, data of depth, T_{max}, S_1/TOC, and HI are considered as inputs of constructed models, and VR is the output. Data normalizing, which improves the performance of constructed models, has been done through the below equation:
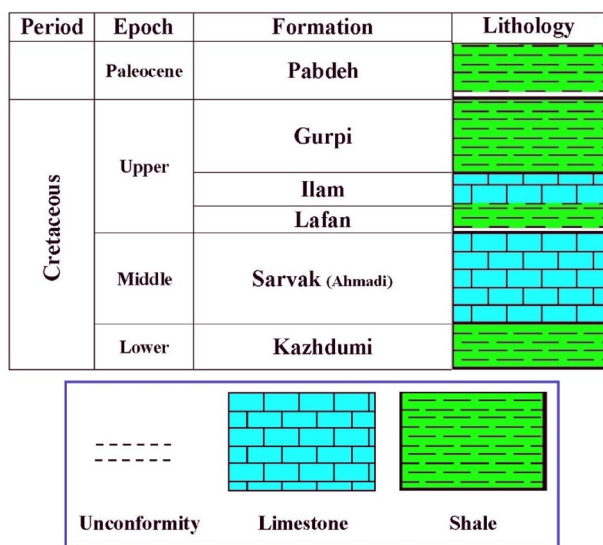
Table 1 Minimum, maximum, and average values of selected inputs and output data

	Parameter	Minimum	Maximum	Average
X_1	Depth (m)	1159	3603	2290.16
X_2	T-max (°C)	410	464	430.51
X_3	S_1/TOC	0.07	1.26	0.48
X_4	HI	84	786	237.74
Y	Ro (%)	0.35	0.82	0.57

$$nz_i = 2 \times \frac{z_i - z_{min}}{z_{max} - z_{min}} - 1, \tag{1}$$

where nz_i refers to the normalized value, i is the number of parameters, z_{min} is the minimum value of z series, and z_{max} refers to the maximum value of them.

Machine learning systems methodology

Machine learning methods are currently excessively used in various fields of science and industry. They are straightforward and flexible tools for accurate prediction. Also, they don't consume so much time for modeling (Lary et al. 2015; Mohaghegh 2017; Al-Fatlawi 2018). These aforementioned factors result in their complete application, especially in oil industries (Anifowose et al. 2017; Kahani et al. 2018; Sabah et al. 2019; Ghaffarkhah et al. 2019; Amin et al. 2019; Cheshmeh Sefidi and Ajorkaran 2019; Esfandiarian et al. 2019). Despite abundant researches based on soft computing methods, no studies have been done in VR prediction through Rock-Eval data, and this article is considered as the first attempt for this purpose which applies 15 distinct intelligent methods of decision tree (DT), support vector machine (SVM), group method of data handling (GMDH), radial basis function (RBF), random forest (RF), multilayer perceptron (MLP), as well as MLP + GOA, MLP + BAT, MLP + PSO, and MLP + GA as well as adaptive neuro-fuzzy inference system (ANFIS), ANFIS + GOA, ANFIS + BAT, ANFIS + PSO, and ANFIS + GA.

Decision tree as supervised learning with a technique of classification and regression tree or CART consists of branches and nodes in a graphical form that is considered as a predictive model in data mining (JiaWei and Micheline 2001; Taylor 2019). The flowchart of the decision tree is illustrated in Fig. 4.

The support vector machine is one of the artificial neural network branches that can be applied for regression and classification purposes through a theory of statistical learning. This approach is also considered as an efficient tool in machine learning as well as data mining (Vapnik 2013; Abbas et al. 2019). In this paper, the kernel function and
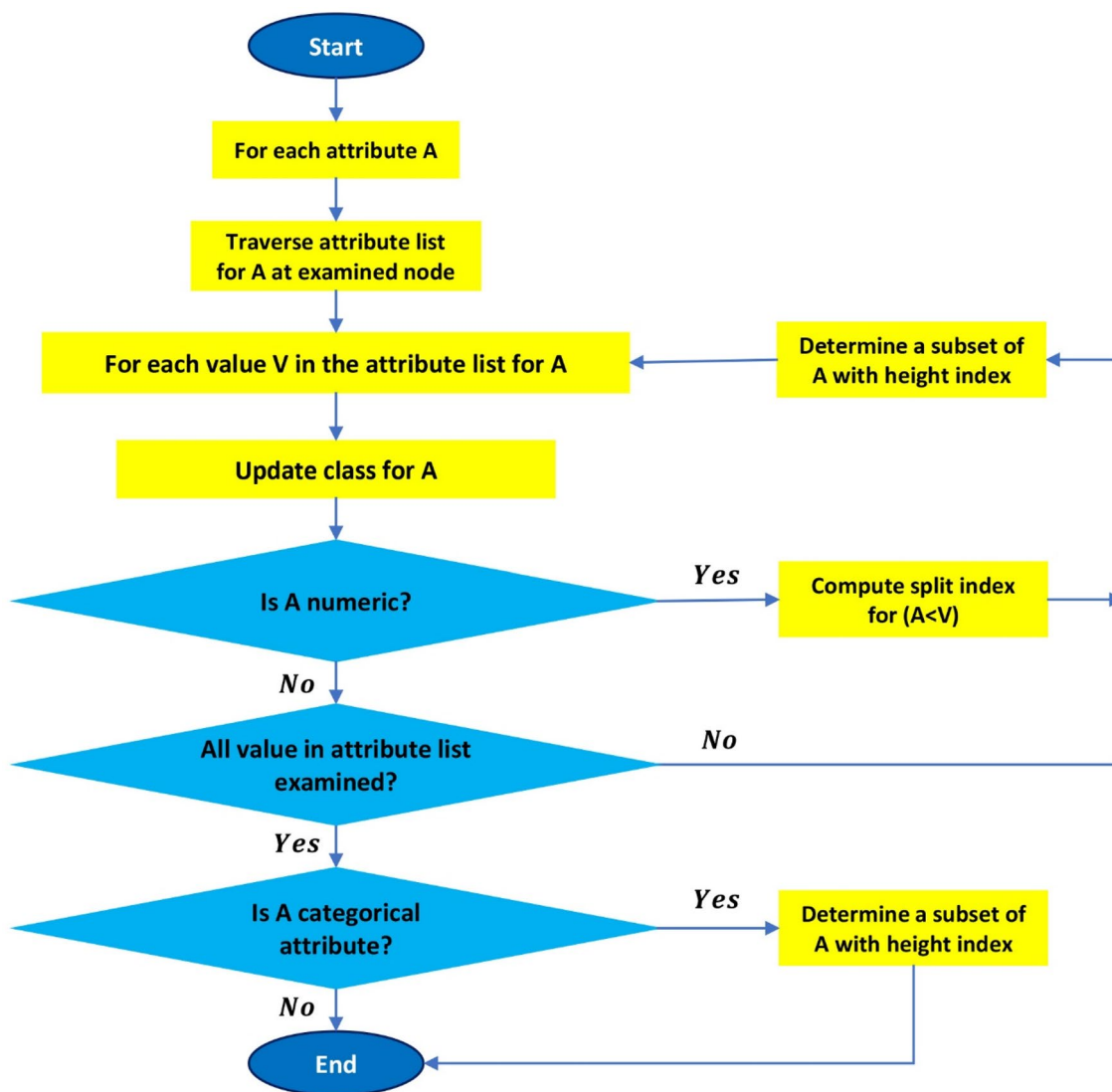

Fig. 3 Simplified stratigraphic column of studied formations

Fig. 4 Schematic of DT workflow

solver of this method are Gaussian and L1QP, respectively. A simple schematic form of SVM is shown in Fig. 5.

Group method of data handling as one of the other applied approaches is a robust connecting algorithm between inputs and outputs, which is proper for settling down non-linear situations as a self-organized system (Hwang 2006; Loni et al. 2018). Detail information on GMDH belongs to this study, and the flowchart is presented in Table 2 and Fig. 6, respectively.

Radial basis functions are types of artificial networks with a structure that employs RBF instead of prevalent activation functions (Amedi et al. 2016). The properties of the operated RBF network are summarized in Table 3.

Moreover, the simplified structure of the RBF network for prediction vitrinite reflectance values as well as its flowchart is illustrated in Figs. 7 and 8, respectively.

One of the significant statistical methods is the random forest, which operates based on the variance concept and can be applied for prediction and sensitivity analysis goals. Moreover, the great advantage of a random forest statistical learning tool is its rapidity (Bosch et al. 2007; Genuer et al. 2010; Aulia et al. 2019). The flowchart of this method is shown in Fig. 9.

Feedforward networks of multilayer perceptions are well-known artificial neural networks with at least three main layers in their structure. In this paper, the MLP network with
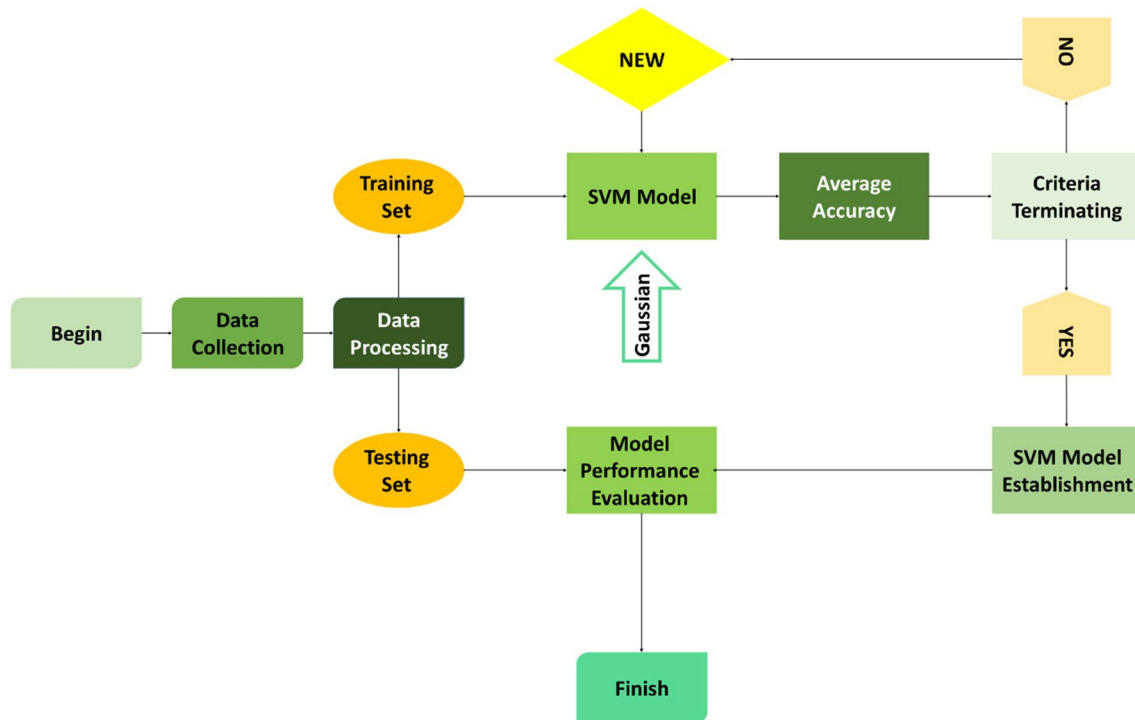
Fig. 5 Schematic of SVM workflow

Table 2 Values of GMDH parameters

GMDH Parameters	Value
Maximum number of neurons in a layer	5
Maximum number of layers	5
Selection pressure	0.6
Train ratio	0.7

2 hidden layers was selected. As is evident, the efficiency of these types of networks depends on the number of neurons in each layer. To that end, several MLP networks with different numbers of neurons were constructed, and then their performances were analyzed by mean square error. This procedure is illustrated in Fig. 10. As shown, the structure with three neurons in the first hidden layer and four neurons in the second hidden layer can be considered as the best situation. Since determining the optimum values of MLP parameters requires several calculations, which takes a lot of time, for increasing efficiency, other parameters are derived from the previous study (Sabah et al. 2019a). Also, the MLP flowchart is presented in Fig. 11. The characteristics of the used MLP network with two hidden layers are presented in Table 4.

The smart hybrid system of ANFIS is the conflation of fuzzy logic and artificial neural network, and this feature makes ANFIS an effective tool and powerful strategy with high accuracy and less time for goals of systems recognition, time series prediction, function approximation, simulation of nonlinear systems, etc. (Jang 1993; Mir et al. 2018). Additional information on ANFIS parameters is listed in Table 5. It must be noted that ANFIS parameters have been determined through previous research (Sabah et al. 2019a) and trial and error approach. Furthermore, the algorithm and the flowchart's architecture are shown in Figs. 12 and 13, respectively.

In addition to the aforementioned methods and to improve the precision of VR prediction, GOA, BAT, GA, and PSO algorithms were used for training MLP neural networks and coupled with ANFIS. Their flowcharts are illustrated in Figs. 14, 15, 16, and 17, respectively. As the name of the grasshopper optimization algorithm implies, this algorithm is derived from the cumulative life of grasshopper in nature. The location of grasshoppers among their swarm can represent a solution for the current issue. Saremi et al. (2017) were the first one who proposed its algorithm relationships, which simulate grasshopper motions. Optimum GOA parameters had been determined by both previous
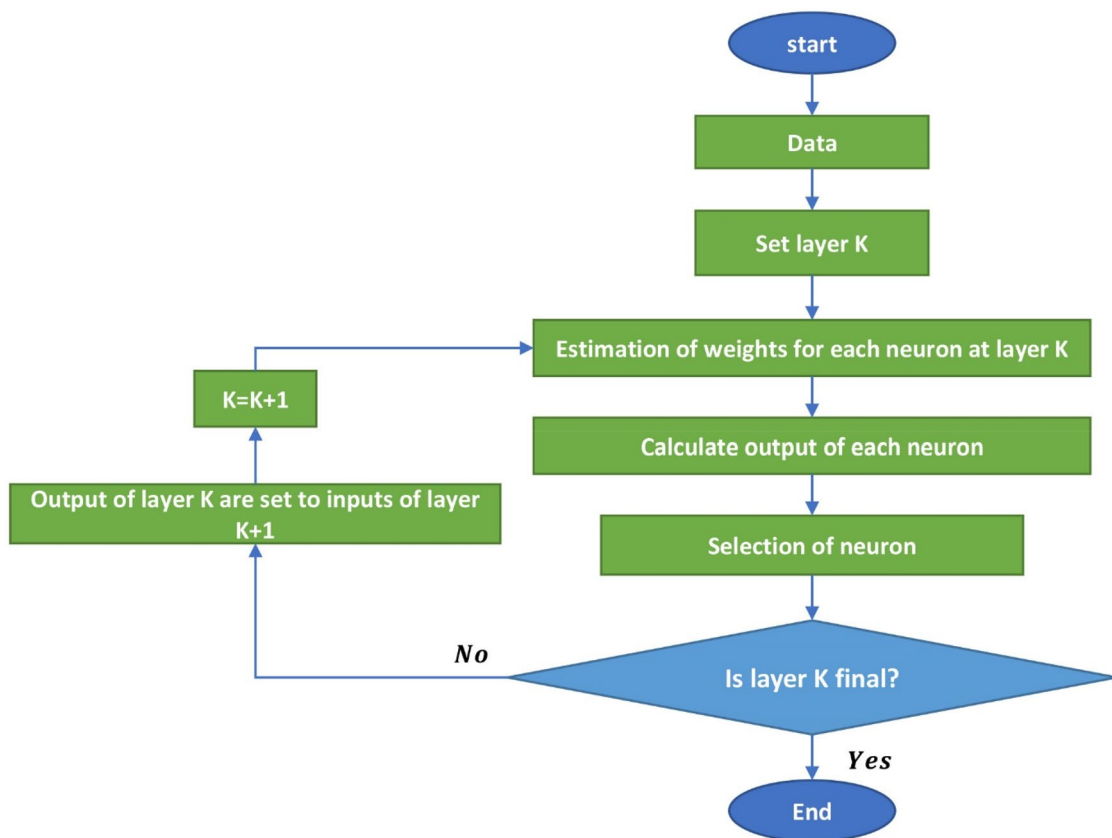
Fig. 6 Schematic of GMDH workflow

Table 3 Properties of RBF network

Network type	Radial basis network
Training function code	newrb
Max neuron number	10
Performance function	MSE
Spread	1
Goal	0

research (Ghaffarkhah et al. 2020) and the trial and error approach. BAT is another algorithm that is inspired by nature. Yang (2010) proposed this optimization algorithm based on bats features for finding bait or way by sending audio pulses. It must be emphasized that this algorithm is quite capable of solving complicated problems (Yang 2010, 2012; Yang and Hossein Gandomi 2012). A particle swarm optimization algorithm as an appropriate tool for searching in a multidimensional space is extremely applied as an optimizer (Atashnezhad et al. 2014). Also, the genetic algorithm is a proper technique for stochastic searching that can find the optimum value of the function by its various processes (Joshi et al. 2006). The control parameters of these four algorithms are completely presented in Table 6.

Results and discussion

As mentioned before, four input values (depth, T_{max}, S_1/TOC, and HI) were used for constructing models to predict vitrinite reflectance values. Pair plot of inputs and output, which provides an appropriate insight into the relationship between all implemented inputs and output values, is illustrated in Fig. 18. This plot reveals that there is a positive correlation between vitrinite reflectance and applied inputs. Moreover, it can be comprehended that VR values are strongly dependent on depth and T_{max}.
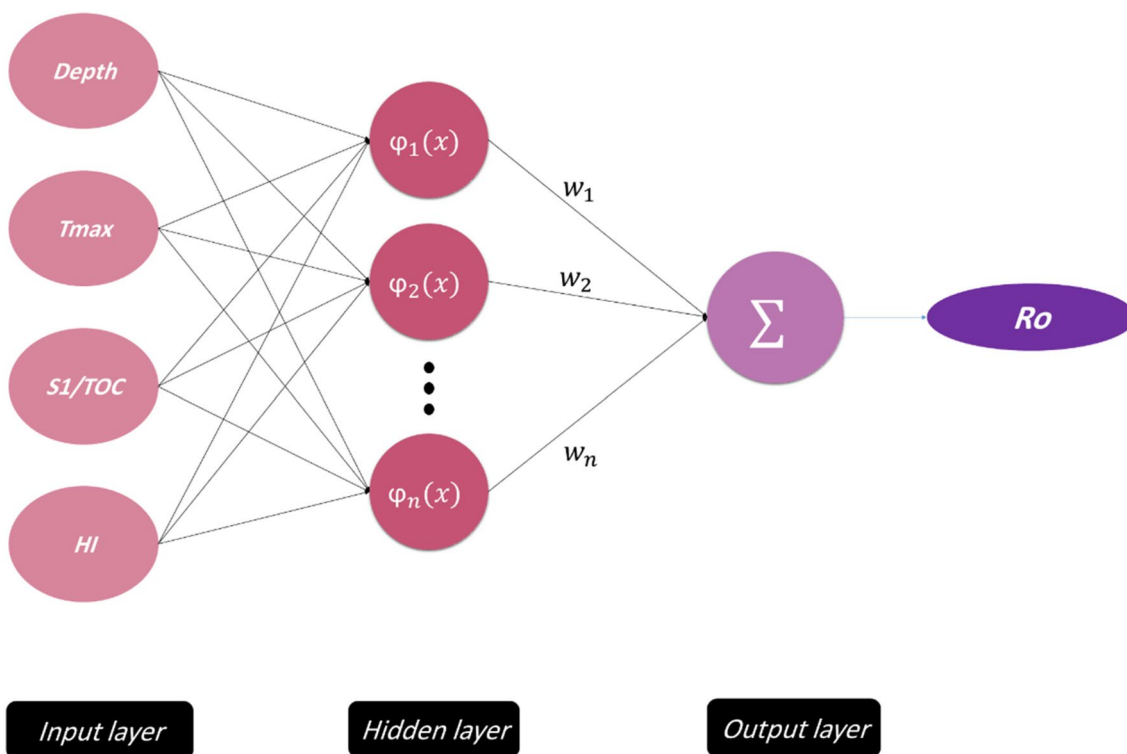
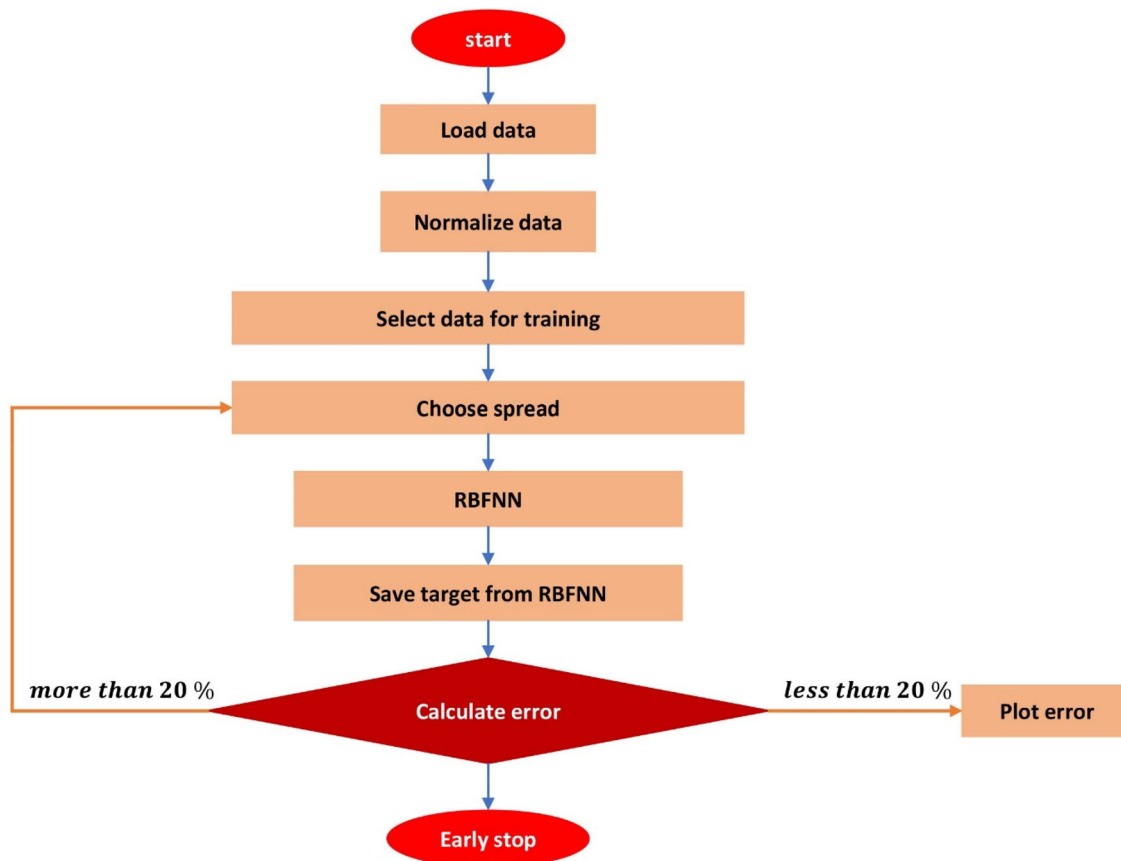**Fig. 7** Schematic form of RBF network structure

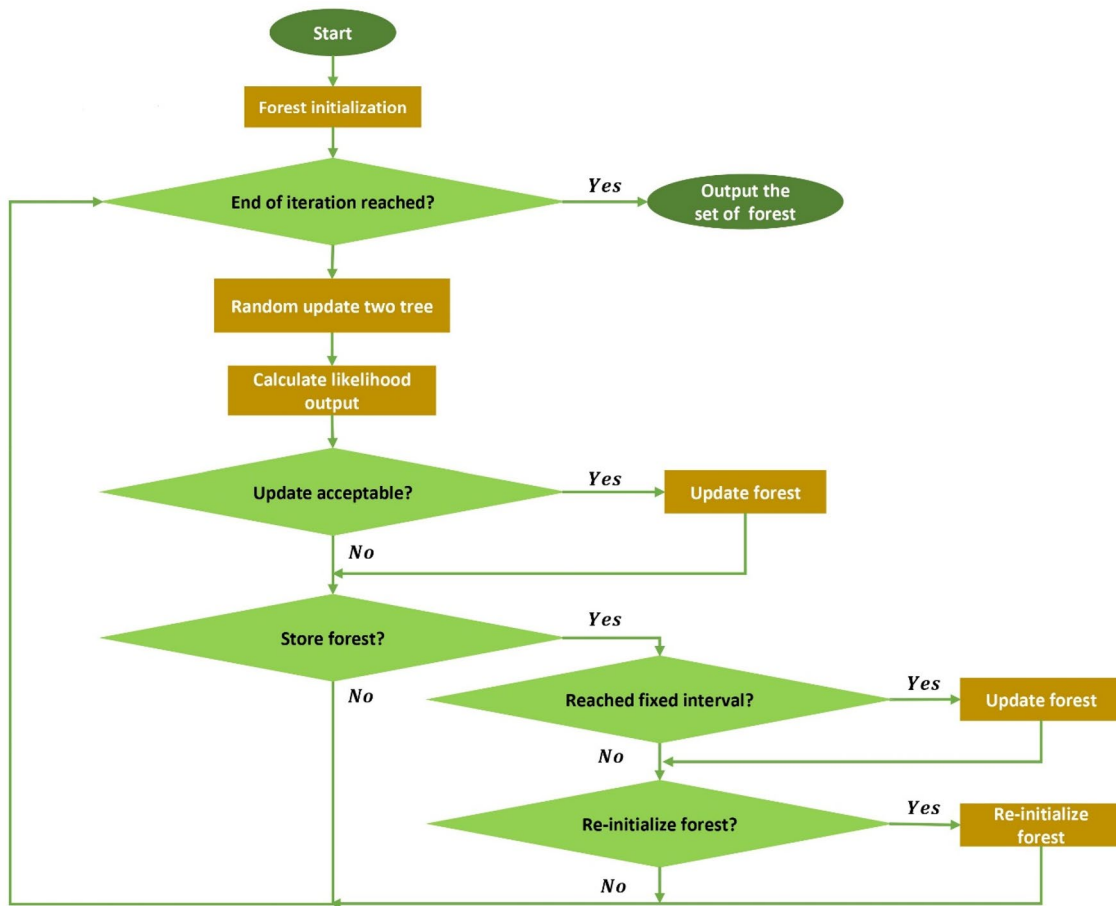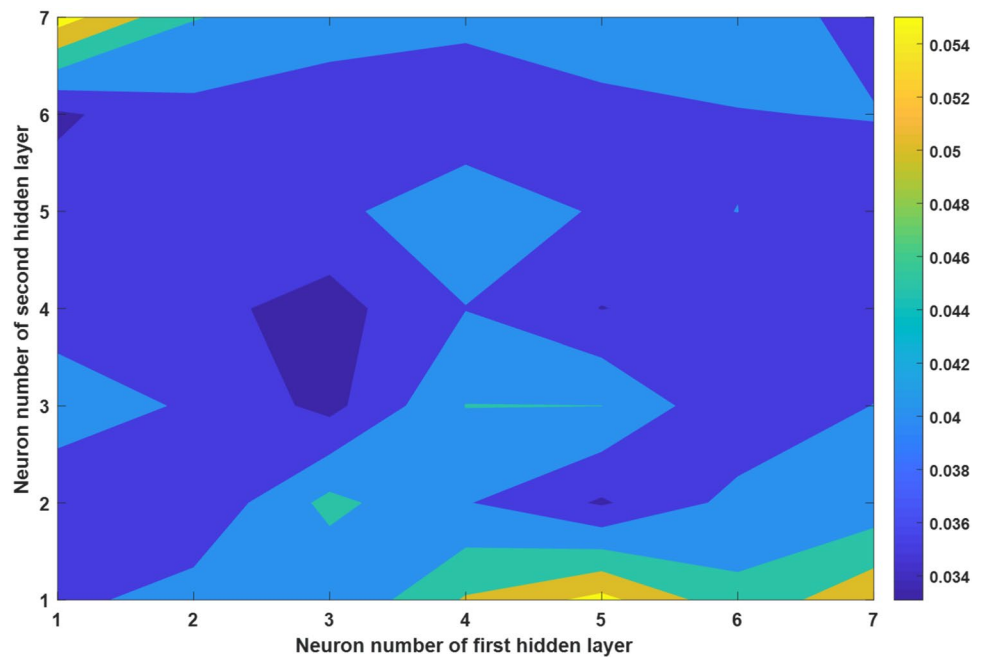**Fig. 8** Schematic of RBF workflow

Fig. 9 Schematic of RF workflow

Fig. 10 Values of MSE for various numbers of neurons in hidden layers of MLP network
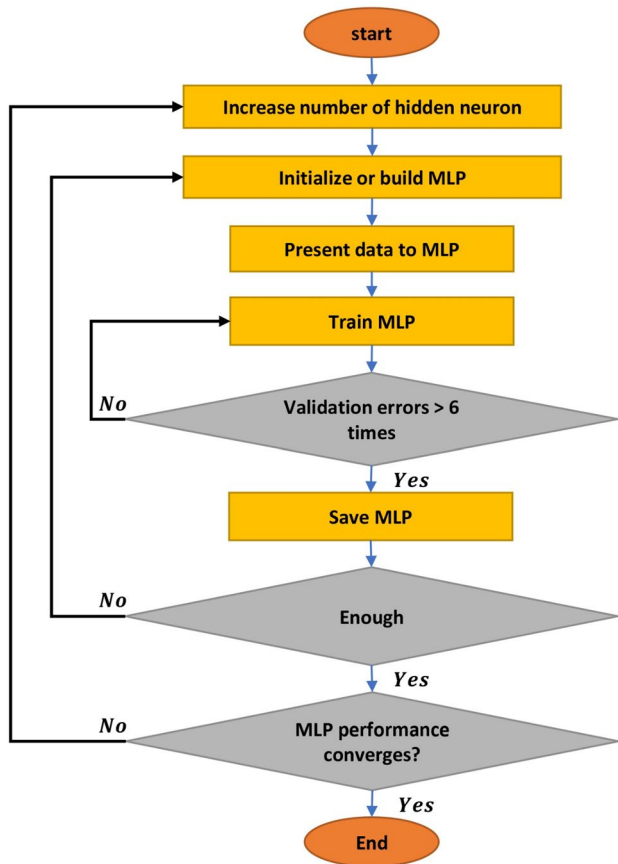
Fig. 11 Schematic of MLP workflow

Table 5 Values of ANFIS parameters

ANFIS Parameters	Value
Number of MFs	5
Radius of influence	0.5
Squash factor	0.5
Accept ratio	0.15
And method	Prod
Or method	Probor
Implication	Min
Aggregation	Max
Defuzzification	Wtaver

Table 4 Architecture parameters of MLP network

| Network type	Multilayer perceptron
Training function	Levenberg–Marquardt backpropagation
Number of layers	3
Nodes in 1st hidden layer	3
Transfer function of 1st hidden layer	TANSIG
Nodes in 2nd hidden layer	4
Transfer function of 2nd hidden layer	TANSIG
Neurons in output layer	1
Transfer function of output layer	PURELIN
Performance objective function	MSE
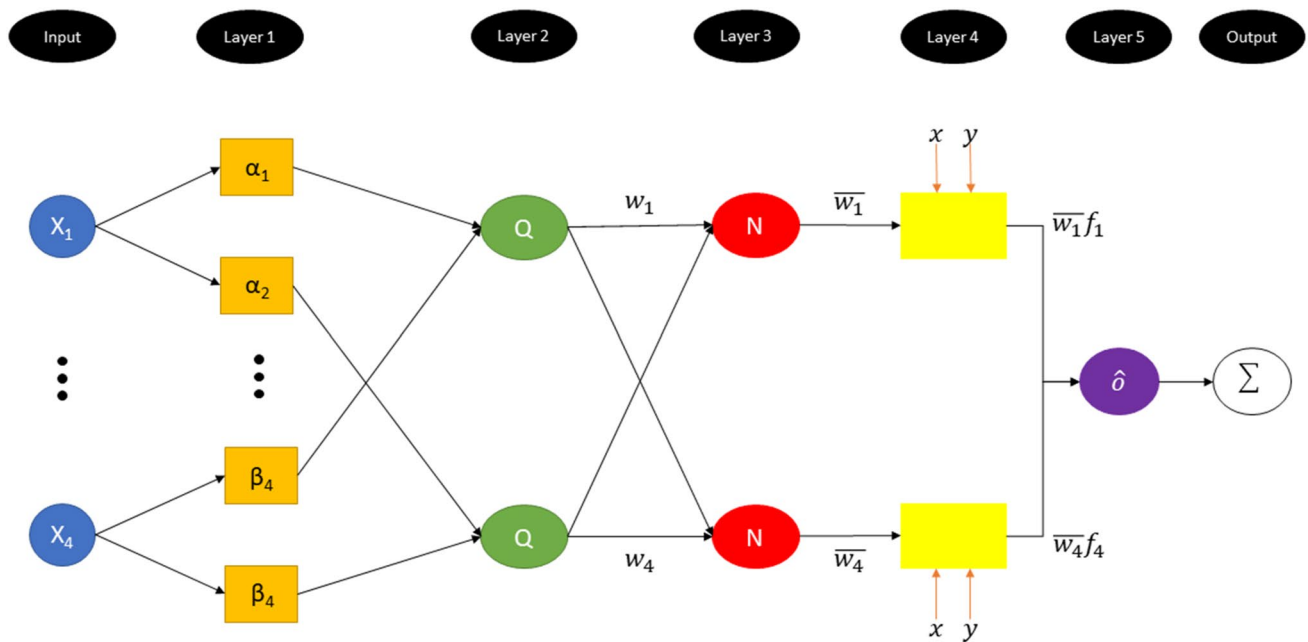
Fig. 12 Simplified structure of the ANFIS algorithm

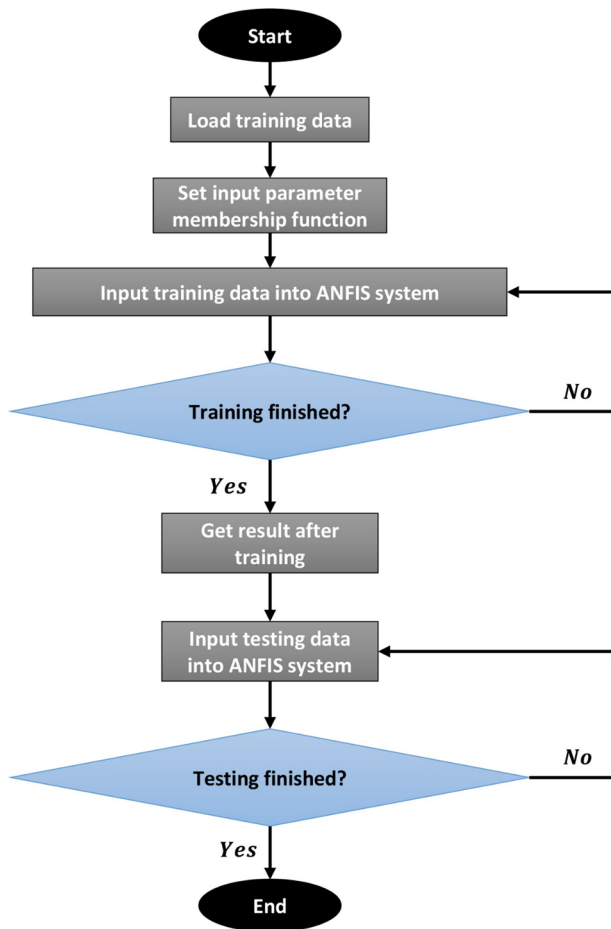**Fig. 13** Schematic of ANFIS workflow

Fig. 14 Schematic of GOA workflow

As a general procedure, constructed models by different aforementioned methods must be compared. To that end, statistical parameters are considered as appropriate tools for appraising the performance of defined models. In this paper, four parameters of average absolute relative deviation (AARD), coefficient of determination (R^2), root mean square error (RMSE), and standard deviation (SD) have been employed for a complete comparison of 15 distinct developed models and selection of the most precise methods for vitrinite reflectance prediction. Equations of these statistical parameters are expressed as follows:

$$\text{AARD} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{O_{i_{\exp}} - O_{i_{\text{pred}}}}{O_{i_{\exp}}} \right| \times 100 \tag{2}$$
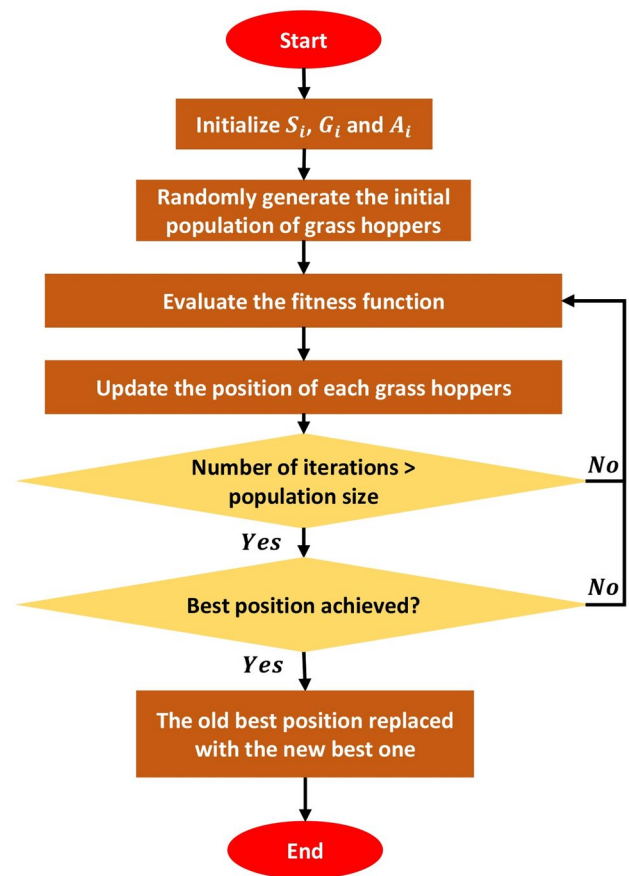
$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(O_{i_{\exp}} - O_{i_{\text{pred}}} \right)^2}{\sum_{i=1}^{N} \left(O_{i_{\text{pred}}} - \overline{O_{\exp}} \right)^2} \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(O_{i_{\exp}} - O_{i_{\text{pred}}} \right)^2} \tag{4}$$

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{O_{i_{\exp}} - O_{i_{\text{pred}}}}{O_{i_{\exp}}} \right)^2}. \tag{5}$$

Four above indicators are calculated for all implemented models, and graphical results for three categories of train data, test data, and total data are illustrated in Figs. 19, 20,
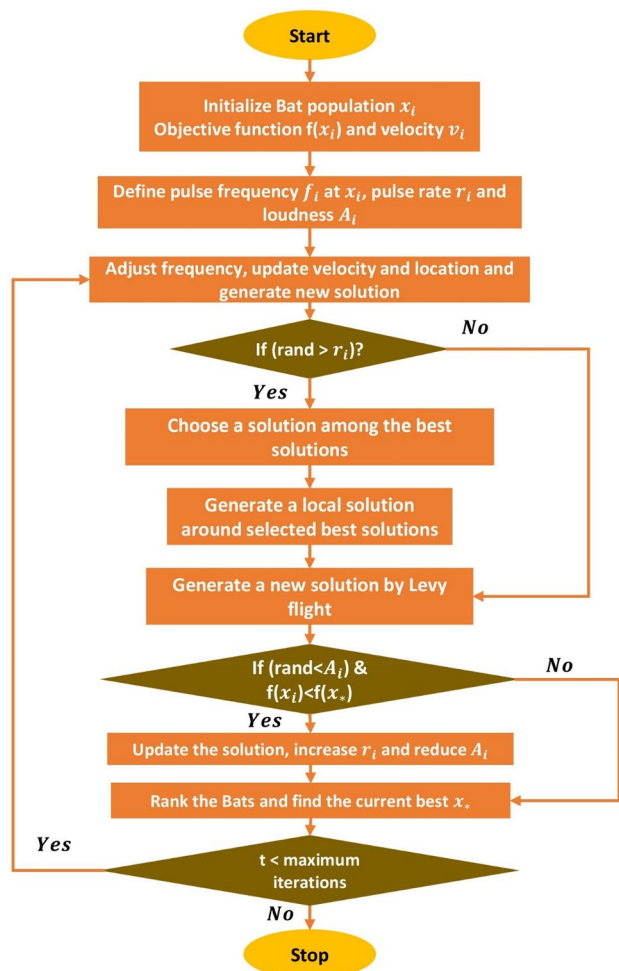
Fig. 15 Schematic of BAT workflow

Fig. 16 Schematic of GA workflow

21, and 22. Figure 19 shows the AARD values of all applied models. As seen, the lowest amount of AARD belongs to the DT model in train and total sets of data, whereas the MLP possesses the lowest AARD in test data. In test data, after MLP, the lowest value belongs to the DT model. The next figure illustrates the R^2 values of the three sets of data. As shown, the highest amount of R^2 is for the decision tree model in all 3 groups of the train, test, and total data (Fig. 20). The variation of RMSE values is shown in Fig. 21. The DT model has the lowest amounts of RMSE. Also, Fig. 22 shows the DT model has the lowest SD values. For better understanding, all these aforementioned data are summarized in Table 7. As can be seen, almost all models
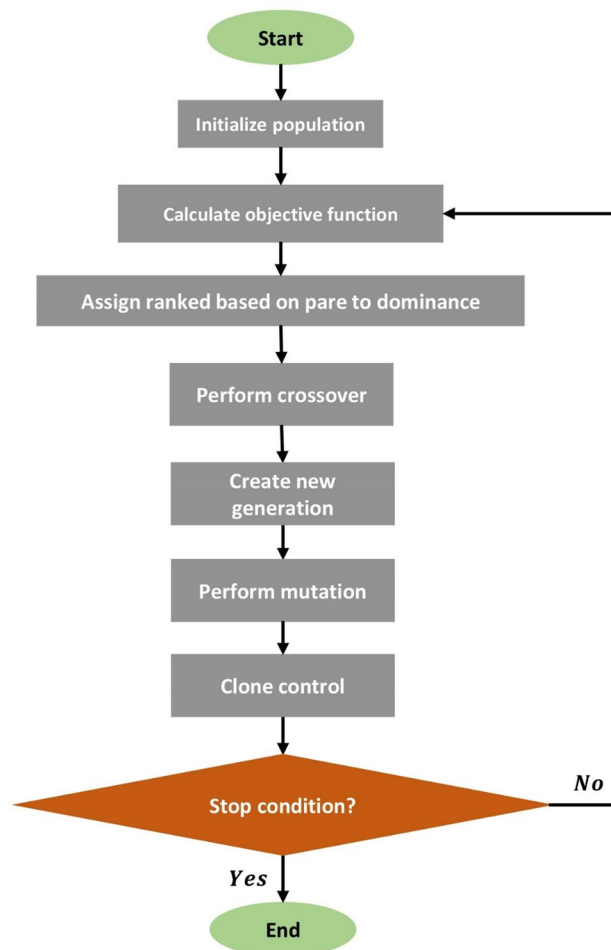
are acting correctly, and the prediction of vitrinite reflectance has been made with high accuracy. However, with a little more precision, it can be realized that among all methods, the decision tree has the best performance, and hence, diagrams of this method merely are presented in the following. The constructed tree for vitrinite reflectance prediction associated with statistical information is shown in Fig. 23. DTs split predictors and form subgroups of separate observations. This process is binary recursive partitioning that divides parent nodes into child nodes (binary splitting). This process is continued to reach terminal nodes that do not have any splitting (Singh 2017). Figure 23 shows the best tree for predicting vitrinite reflectance after running the decision tree model. As illustrated, this tree consists of nodes and leaves.
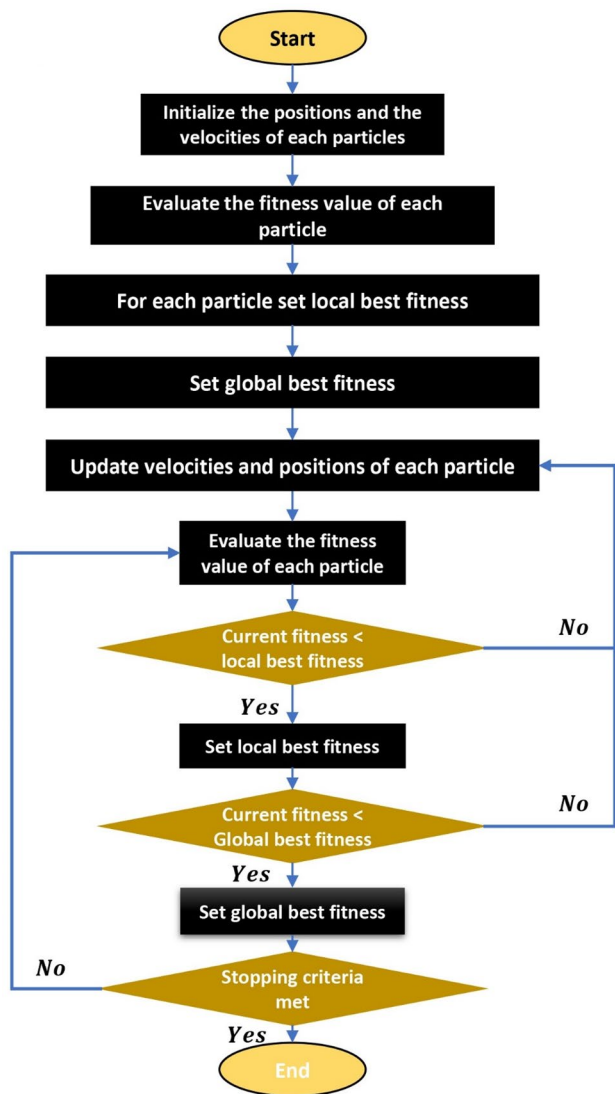
Fig. 17 Schematic of PSO workflow

Table 6 Parameters of GOA, BAT, GA, and PSO

	Value
GOA	
GOA property	
Population size	50
Max iteration	100
BAT	
BAT property	
Swarm size	50
Max generation	100
Loudness	0.5
Pulse rate	0.5
GA	
GA property	
Population size	150
Max generation	100
Selection mode	Tournament
Recommendation percent	40
Crossover percent	50
Mutation percent	10
PSO	
PSO property	
Swarm size	150
Max iteration	100
Cognition coefficient	2
Social coefficient	2

This tree takes the inputs and gives vitrinite reflectance values in leaves. The first input is T_{max}. If the T_{max} value is less than 432.5 °C, the tree chooses the left side to continue the route, and if T_{max} is more than this amount, the right path will be selected. In the next steps, depth, T_{max}, and S1/TOC are determinant factors to continue. Final results of predicted vitrinite reflectance values are presented in the leaves of the tree.

Furthermore, the relative deviation plot of the decision tree method is depicted in Fig. 24. This figure indicates that there is a small deviation between experimental data of vitrinite reflectance and predicted values by the DT method. Obviously, this slight difference suggests the accuracy of this method. To provide a complete sense of the precision of the decision tree model for vitrinite reflectance prediction, the plot of experimental and predicted data is presented in Fig. 25. As it is obvious, DT has been able to predict Ro values with high accuracy, and measured and predicted data are close to each other. The correlation of experimental and predicted data for both train and test data is illustrated in Fig. 26. This figure depicts the DT method as very precise in training data and the test dataset.

In the next step, the achievement of this paper should be compared with previous experimental equations. As mentioned, four equations had been presented for vitrinite reflectance prediction based on T_{max} data. These formulas are as follows:

$$R_o = (1/30)T_{max} - 13.5 \quad \text{(Galimov and Rabbani 2001)} \tag{6}$$

$$R_o = (0.0180)T_{max} - 7.16 \quad \text{(Jarvie et al.2001)} \tag{7}$$

$$R_o = (0.0085)T_{max} - 2.7914 \quad \text{(Wust et al.2013)} \tag{8}$$

$$R_o = (0.0149)T_{max} - 5.8593 \quad \text{(Wust et al.2013).} \tag{9}$$
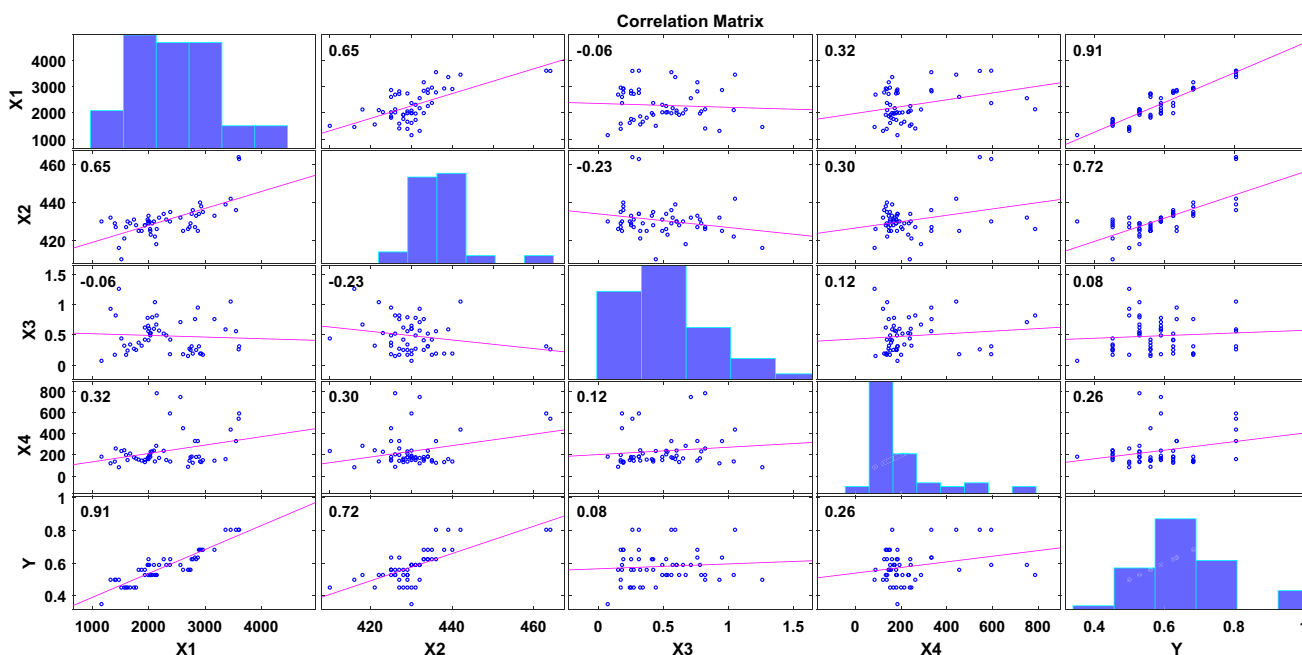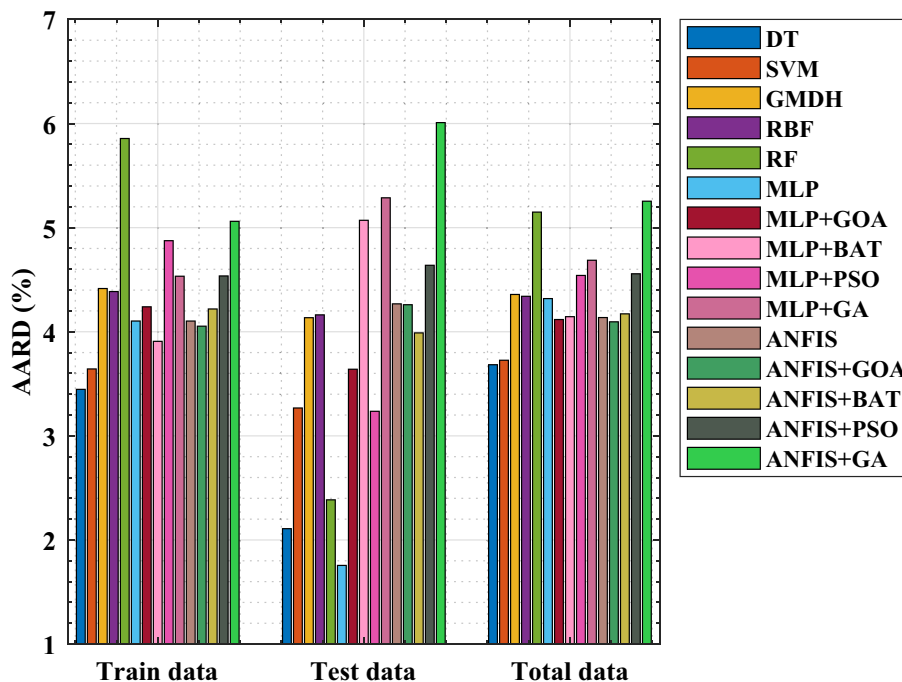
Correlation Matrix

Fig. 18 The plot of the relationship between inputs and output values

Fig. 19 Comparison of developed models on the basis of the AARD statistical parameter

To better understand the performance of methods based on experimental equations and artificial intelligence methods, the vitrinite reflectance values have been calculated by Eqs. 6–9 based on this paper's presented T_{max} data. The statistical parameters (AARD, R^2, RMSE, SD) are computed for these experimental equations and summarized in Table 8. For better understanding, the statistical parameters of the decision tree are also listed in this table. As seen, the DT method with the lowest amounts of AARD, RMSE, SD, and the highest value of R^2 is the most

Fig. 20 Comparison of developed models on the basis of the R^2 statistical parameter
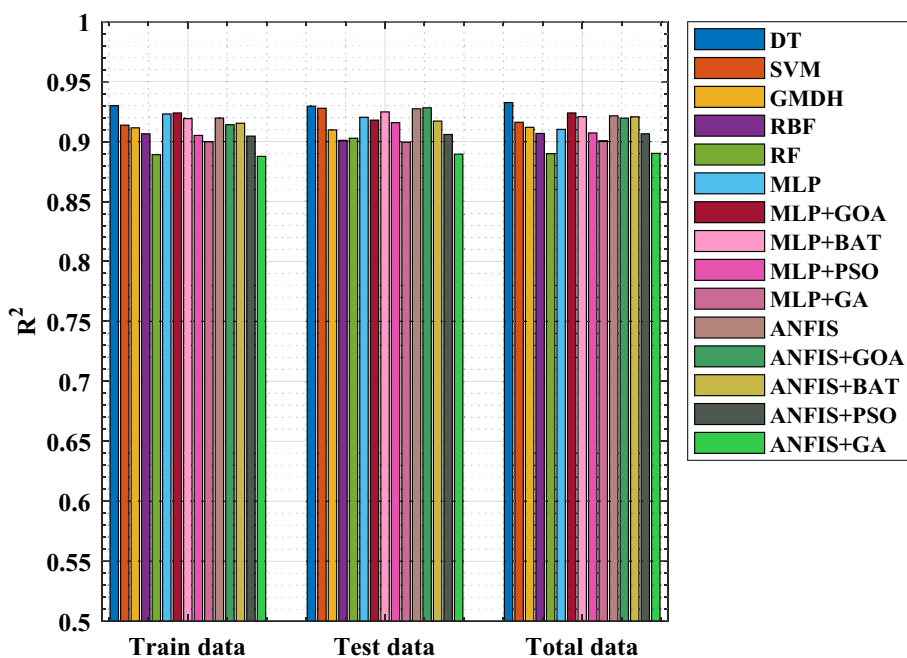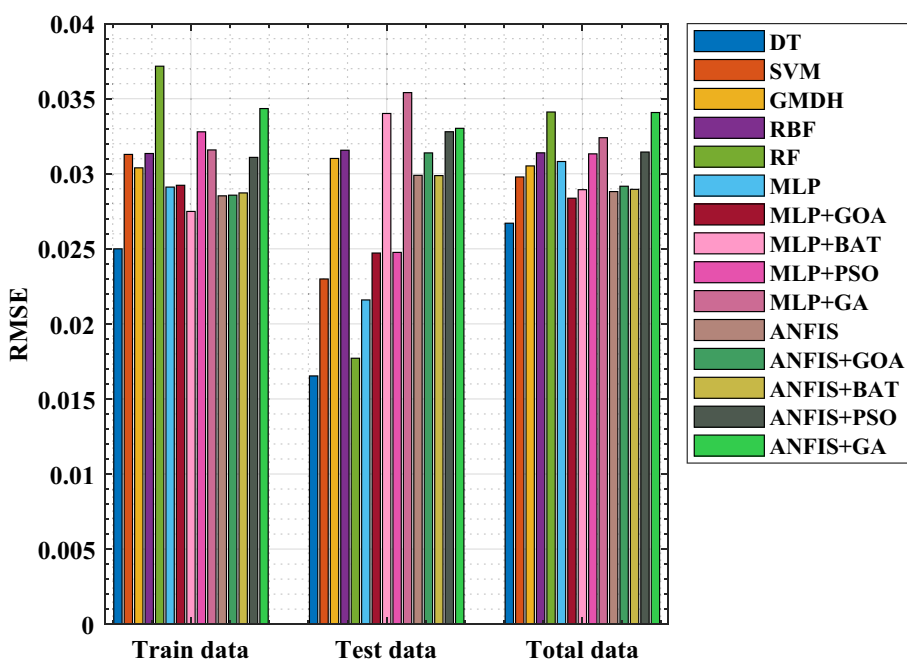

Fig. 21 Comparison of developed models on the basis of the RMSE statistical parameter

precise method for vitrinite reflectance prediction. The results of Table 8 show the poor performance of experimental equations. These results indicate the high accuracy of machine learning approaches compared to experimental methods. The variations of predicted vitrinite reflectance quantities by empirical equations and the decision tree approach (as the most precise machine learning method) with depth are illustrated in Fig. 27. Empirical equations perform weakly for vitrinite reflectance prediction.

Fig. 22 Comparison of developed models on the basis of the SD statistical parameter
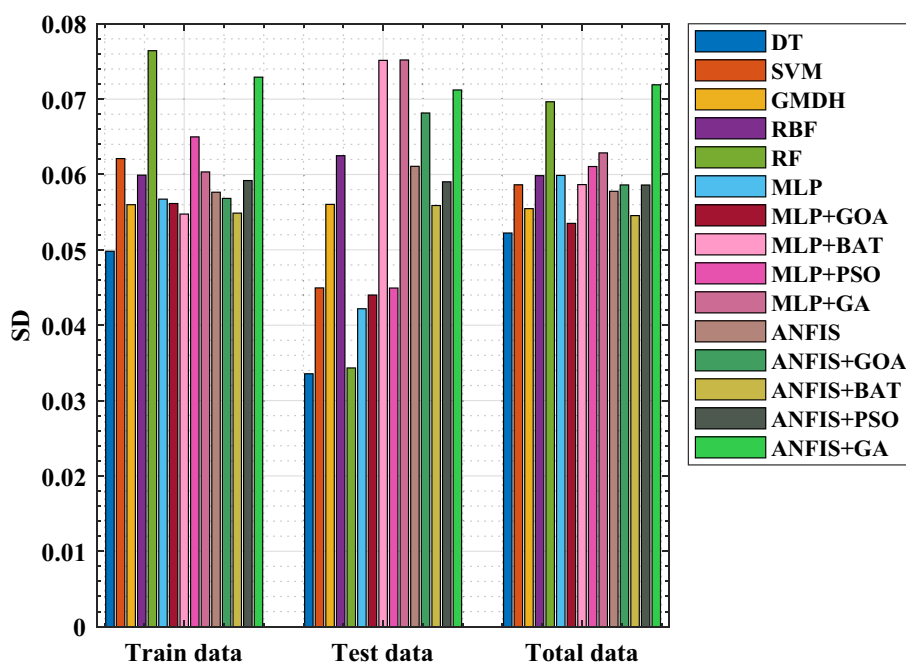
Table 7 Statistical parameters values for all constructed models

	Training				Test				Overall			
	RMSE	AARD (%)	R^2	SD	RMSE	AARD (%)	R^2	SD	RMSE	AARD (%)	R^2	SD
DT	0.0249	3.44	0.929	0.04	0.0165	2.10	0.929	0.03	0.0267	3.68	0.932	0.05
SVM	0.0312	3.64	0.913	0.06	0.0229	3.26	0.927	0.04	0.0297	3.72	0.916	0.05
GMDH	0.030	4.41	0.911	0.05	0.0310	4.13	0.909	0.05	0.030	4.35	0.911	0.05
RBF	0.031	4.38	0.906	0.05	0.0315	4.16	0.901	0.06	0.031	4.34	0.906	0.05
RF	0.037	5.85	0.889	0.07	0.0177	2.38	0.902	0.03	0.034	5.14	0.889	0.06
MLP	0.029	4.10	0.923	0.05	0.0215	1.75	0.920	0.04	0.030	4.31	0.910	0.05
MLP + GOA	0.029	4.23	0.923	0.05	0.0247	3.63	0.917	0.04	0.028	4.11	0.923	0.05
MLP + BAT	0.027	3.90	0.919	0.05	0.0340	5.07	0.924	0.07	0.028	4.14	0.920	0.05
MLP + PSO	0.032	4.87	0.905	0.06	0.0247	3.23	0.915	0.04	0.031	4.54	0.907	0.06
MLP + GA	0.031	4.53	0.899	0.06	0.0354	5.28	0.899	0.07	0.032	4.68	0.900	0.06
ANFIS	0.028	4.10	0.919	0.05	0.0298	4.26	0.927	0.06	0.028	4.13	0.921	0.05
ANFIS + GOA	0.028	4.05	0.914	0.05	0.0313	4.25	0.928	0.06	0.029	4.09	0.919	0.05
ANFIS + BAT	0.028	4.21	0.915	0.05	0.0298	3.98	0.917	0.05	0.028	4.17	0.920	0.05
ANFIS + PSO	0.031	4.53	0.904	0.05	0.0327	4.63	0.905	0.05	0.031	4.55	0.906	0.05
ANFIS + GA	0.034	5.06	0.887	0.07	0.0330	6.00	0.889	0.07	0.034	4.25	0.890	0.07

Conclusions

Vitrinite reflectance values have been predicted by using machine learning approaches for the first time in this paper. For this purpose, several Rock-Eval parameters were used to construct machine learning models. For this purpose, in addition to depth, T_{max}, S₁/TOC, and

HI values were selected for vitrinite reflectance prediction. Decision tree (DT), support vector machine (SVM), group method of data handling (GMDH), radial basis function (RBF), random forest (RF), multilayer perceptron (MLP), MLP + GOA, MLP + BAT, MLP + PSO and MLP + GA, adaptive neuro-fuzzy inference system (ANFIS), ANFIS + GOA, ANFIS + BAT, ANFIS + PSO,
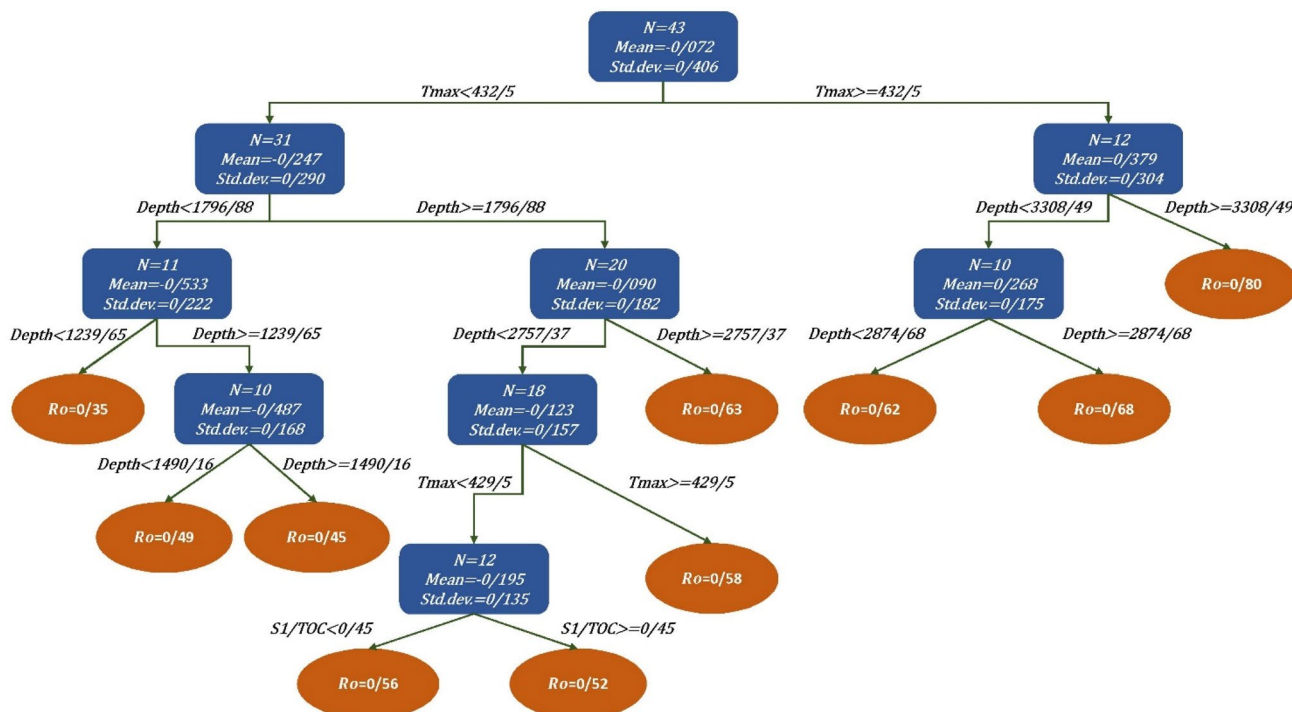
Fig. 23 An implemented tree for vitrinite reflectance prediction. Statistical information and vitrinite reflectance values are shown in blue rectangles and orange ellipsoids, respectively
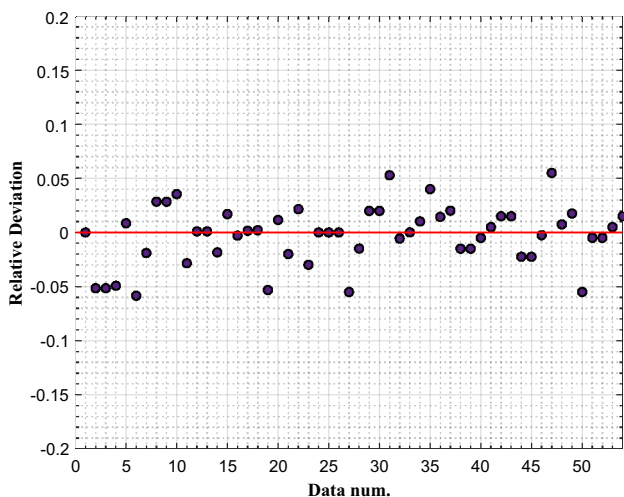

Fig. 24 Relative deviation plot of the decision tree model
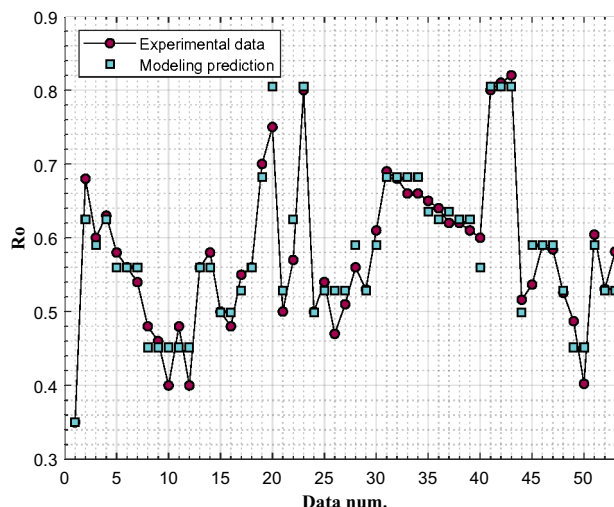

Fig. 25 Experimental data of vitrinite reflectance and prediction values by DT

and ANFIS + GA are 15 implemented methods for vitrinite reflectance prediction. After normalizing data, constructed models were compared via four statistical parameters of average absolute relative deviation (AARD), coefficient of determination (R^2), root mean square error (RMSE), and standard deviation (SD). Results of the comparison indicate that all methods are precise, and their outputs are
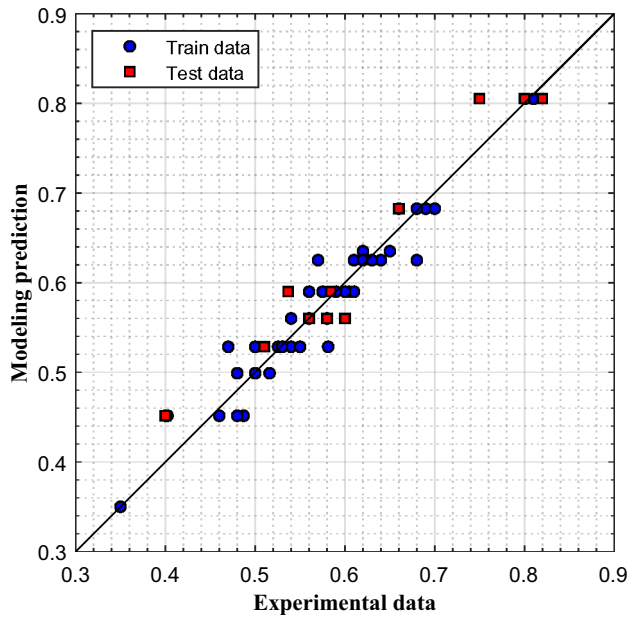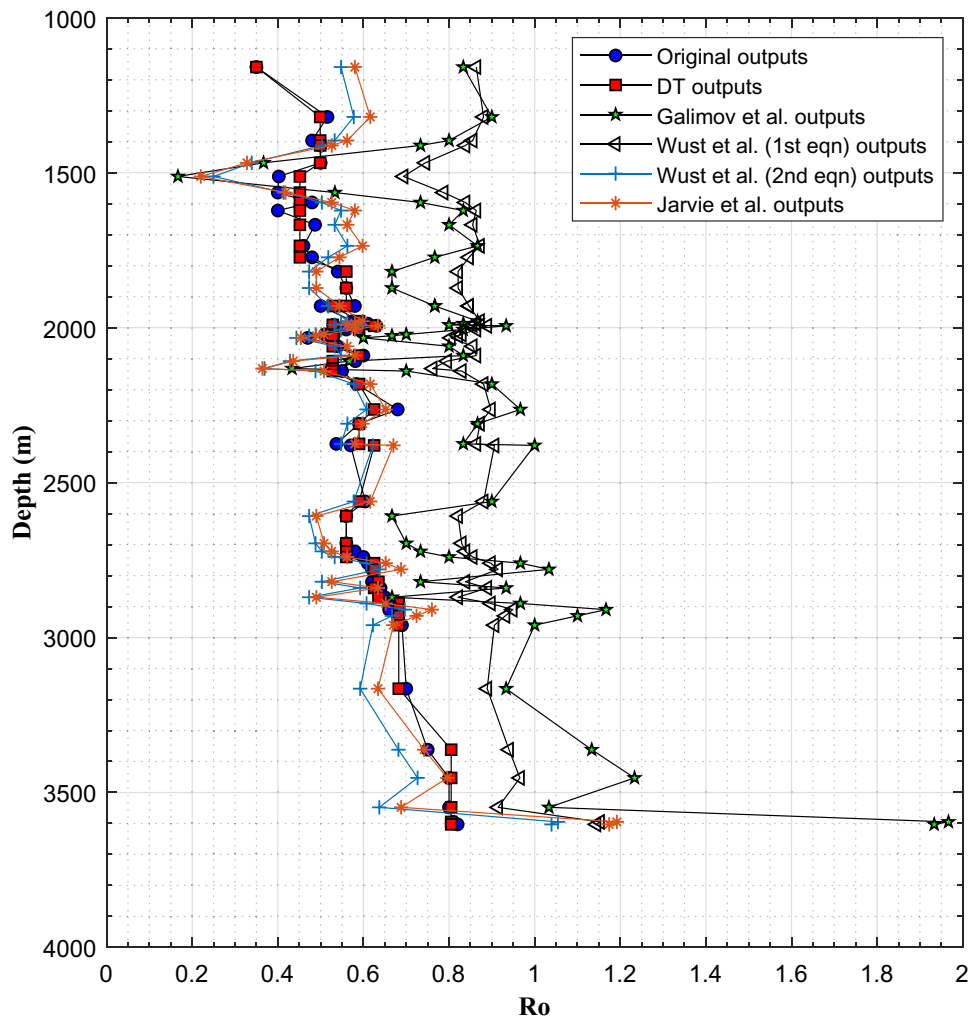
Fig. 26 Predicted values of Ro by DT versus experimental values for both train and test datasets

Table 8 The calculated statistical parameters for DT method and empirical equations of 6–9

	DT	Equation 6	Equation 7	Equation 8	Equation 9
AARD	3.683	50.014	13.307	53.535	12.482
R^2	0.933	− 10.757	− 0.102	− 7.420	0.199
RMSE	0.027	0.353	0.108	0.298	0.092
SD	0.052	0.585	0.196	0.587	0.170

acceptable, whereas the decision tree can be considered as the most accurate method, which possesses the lowest relative deviation between measured and predicted vitrinite reflectance values. A comparison between the decision tree and previously proposed equations for vitrinite reflectance prediction indicates that the machine learning approach performs more accurately than empirical equations.

Fig. 27 The variation of vitrinite reflectance quantities with depth for decision tree and 4 empirical equations. *Note* that 1st eqn and 2nd eqn are Eqs. 8 and 9, respectively

References

Abbas AK, Al-haideri NA, Bashikh AA (2019) Implementing artificial neural networks and support vector machines to predict lost circulation. Egypt J Pet. https://doi.org/10.1016/j.ejpe.2019.06.006

Al-Fatlawi OF (2018) Numerical simulation for the reserve estimation and production optimization from tight gas reservoirs. PhD Thesis, Curtin University of Technology

Amedi HR, Baghban A, Ahmadi MA (2016) Evolving machine learning models to predict hydrogen sulfide solubility in the presence of various ionic liquids. J Mol Liq 216:411–422. https://doi.org/10.1016/j.molliq.2016.01.060

Amin JS, Kuyakhi HR, Bahadori A (2019) 'Intelligent prediction of aliphatic and aromatic hydrocarbons in Caspian Sea sediment using a neural network based on particle swarm optimization. Pet Sc Technol. https://doi.org/10.1080/10916466.2018.1542439

Anifowose FA, Labadin J, Abdulraheem A (2017) Ensemble machine learning: an untapped modeling paradigm for petroleum reservoir characterization. J Pet Sci Eng 151:480–487

Atashnezhad A, Wood DA, Fereidounpour A, Khosravanian R (2014) Designing and optimizing deviated wellbore trajectories using novel particle swarm algorithms. J Nat Gas Sci Eng 21:1184–1204

Aulia A, Jeong D, Saaid IM, Kania D, Shuker MT, El-Khatib NA (2019) A Random Forests-based sensitivity analysis framework for assisted history matching. JPet Sci Eng. https://doi.org/10.1016/j.petrol.2019.106237

Barker CE, Elders WA (1981) Vitrinite reflectance geothermometry and apparent heating duration in the Cerro Prieto geothermal field. Geothermics 10(3–4):207–223

Behar F, Beaumont V, Penteado HLDB (2001) Rock-Eval 6 technology: performances and developments. Oil Gas Sci Technol 56(2):111–134

Bertrand R (1990) Correlations among the reflectances of vitrinite, chitinozoans, graptolites and scolecodonts. Org Geochem 15(6):565–574

Bosch A, Zisserman A, Munoz X (2007) Image classification using random forests and ferns. In: IEEE 11th international conference on computer vision, 14–21: 1–8. https://doi.org/10.1109/ICCV.2007.4409066

Chen W, Zhang Q, Zhang C, Dai S, Chen L, Wang C, Kang W, Wang J, Jiao S (2017) Marine Vitrinite: Vitrinite Reflectance as an Indicator of Concealed Pluton. Geotectonica et Metallogenia 41

Cheshire S, Craddock PR, Xu G, Sauerer B, Pomerantz AE, McCormick D, Abdallah W (2017) Assessing thermal maturity beyond the reaches of vitrinite reflectance and rock-eval pyrolysis: a case study from the Silurian Qusaiba formation. Int J Coal Geol 180:29–45. https://doi.org/10.1016/j.coal.2017.07.006

Cheshmeh-Sefidi A, Ajorkaran F (2019) A novel MLP-ANN approach to predict solution gas-oil ratio. Pet Sci Technol 37(23):2302–2308. https://doi.org/10.1080/10916466.2018.1490759

Cole GA (1994) Graptolite-chitinozoan reflectance and its relationship to other geochemical maturity indicators in the Silurian Qusaiba Shale, Saudi Arabia. Energy Fuels 8(6):1443–1459

Curiale JA (1986) Origin of solid bitumens, with emphasis on biological marker results. Org Geochem 10(1–3):559–580

Dembicki H (2016) Practical petroleum geochemistry for exploration and production. Elsevier, Amsterdam

Dembicki H Jr (2009) Three common source rock evaluation errors made by geologists during prospect or play appraisals. AAPG Bull 93(3):341–356

Epstein AG, Epstein JB, Harris LD (1976) Conodont color alteration: an index to organic metamorphism (Geological Survey professional paper; 995). Library of Congress Cataloging in Publication Data, United States Governlvient Printing Office, Washington

Esfandiarian A, Sedaghat M, Maniatpour A, Darvish H (2019) Application of grid partitioning based fuzzy inference system as a novel predictor to estimate dynamic viscosity of n-alkane. Pet Sci Technol 37(23):2309–2314. https://doi.org/10.1080/10916466.2018.1490760

Espitalié J (1986) Use of T_{max} as a maturation index for different types of organic matter comparison with vitrinite reflectance. Therm Model Sediment Basins 44:475–496

Espitalié J, Laporte JL, Madec M, Marquis F, Leplat P, Paulet J, Boutefeu A (1977) Méthode rapide de caractérisation des roches mètres, de leur potentiel pétrolier et de leur degré d'évolution. Revue de l'Institut français du Pétrole 32(1):23–42

Galimov EM, Rabbani AR (2001) Geochemical characteristics and origin of natural gas in southern Iran. Geochem Int 39(8):780–792

Ge X, Wang Y, Fan Y, Fan Z, Deng S (2015) (2015) Determination of total organic carbon (TOC) in tight reservoir using empirical mode decomposition-support vector regression (EMD-SVR): a case study from XX-1 Basin Western China. ASEG Ext Abstr 1:1–10

Genuer R, Poggi J-M, Tuleau-Malot C (2010) Variable selection using random forests. Pattern Recognit Lett 31(14):2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

Ghaffarkhah A, Afrand M, Talebkeikhah M, Sehat AA, Moraveji MK, Talebkeikhah F, Arjmand M (2020) On evaluation of thermophysical properties of transformer oil-based nanofluids: a comprehensive modeling and experimental study. J Mol Liq 300:112249

Ghaffarkhah A, Bazzi A, Dijvejin ZA, Talebkeikhah M, Moraveji MK, Agin F (2019) Experimental and numerical analysis of rheological characterization of hybrid nano-lubricants containing COOH-Functionalized MWCNTs and oxide nanoparticles. Int Commun Heat Mass Transf 101:103–115

Ghazban F (2009) Petroleum geology of the Persian Gulf. Tehran University Press, Tehran

Haghi AH, Kharrat R, Asef MR, Rezazadegan H (2013) Present-day stress of the central Persian Gulf: implications for drilling and well performance. Tectonophysics 608:1429–1441

Homke S, Vergés J, Serra-Kiel J, Bernaola G, Sharp I, Garcés M, Montero-Verdú I, Karpuz R, Goodarzi MH (2009) Late Cretaceous-Paleocene formation of the proto–Zagros foreland basin, Lurestan Province, SW Iran. Geol Soc Am Bull 121(7–8):963–978

Hou M, Zha M, Ding X, Imin A, Lai R, Pan S, Ding Y (2020) A prediction model of vitrinite reflectance for suppression of organic-matter maturation by overpressure. J Geol 128(2):189–200

Hwang HS (2006) Fuzzy GMDH-type neural network model and its application to forecasting of mobile communication. Comput Ind Eng 50(4):450–457. https://doi.org/10.1016/j.cie.2005.08.005

Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685

Jarvie, D. M. (2012) Shale resource systems for oil and gas: Part 2—Shale-oil resource systems

Jarvie DM, Claxton BL, Henk F, Breyer JT (2001) Oil and shale gas from the Barnett shale. Ft. Worth Basin, Texas. Talk presented at the AAPG National Convention, Denver, CO', American Association of Petroleum Geologists Bulletin A. pp 100

Jiang J, Yang W, Cheng Y, Liu Z, Zhang Q, Zhao K (2019) 'Molecular structure characterization of middle-high rank coal via XRD Raman and FTIR spectroscopy: Implications for coalification. Fuel 239:559–572

JiaWei H, Micheline K (2001) Data mining: concepts and techniques. Morgan Kaufmann, Burlington

Johnson LM, Rezaee R, Kadkhodaie A, Smith G, Yu H (2018) Geochemical property modelling of a potential shale reservoir in the Canning Basin (Western Australia), using artificial neural networks and geostatistical tools. Comput Geosci 120:73–81

Joshi D, Sandhu KS, Soni MK (2006) Constant voltage constant frequency operation for a self-excited induction generator. IEEE Trans Energy Convers 21(1):228–234

Kadkhodaie A, Rezaee R (2017) Estimation of vitrinite reflectance from well log data. J Pet Sci Eng 148:94–102

Kahani M, Ahmadi MH, Tatar A, Sadeghzadeh M (2018) Development of multilayer perceptron artificial neural network (MLP-ANN) and least square support vector machine (LSSVM) models to predict Nusselt number and pressure drop of TiO_2/water nanofluid flows through non-straight pathways. Numer Heat Transf Part A Appl 74(4):1190–1206

Katz BJ (1983) Limitations of 'Rock-Eval'pyrolysis for typing organic matter. Org Geochem 4(3–4):195–199

Khoshnoodkia M, Mohseni H, Rahmani O, Mohammadi A (2011) TOC determination of Gadvan formation in South Pars Gas field, using artificial intelligent systems and geochemical data. J Pet Sci Eng 78(1):119–130

Kibria MG, Das S, Hu Q-H, Basu AR, Hu W-X, Mandal S (2020) 'Thermal maturity evaluation using Raman spectroscopy for oil shale samples of USA: comparisons with vitrinite reflectance and pyrolysis methods. Pet Sci 17:567–581

Lary DJ, Alavi AH, Gandomi AH, Walker AL (2015) Machine learning in geosciences and remote sensing. Geosci Front 30:1e9

Liu B, Teng J, Mastalerz M, Schieber J (2020) Assessing the thermal maturity of black shales using vitrinite reflectance: insights from Devonian black shales in the eastern United States. Int J Coal Geol 220:103426. https://doi.org/10.1016/j.coal.2020.103426

Loni R, Asli-Ardeh EA, Ghobadian B, Ahmadi MH, Bellos E (2018) GMDH modeling and experimental investigation of thermal performance enhancement of hemispherical cavity receiver using MWCNT/oil nanofluid. Sol Energy 171:790–803

Lupoi JS, Hackley PC, Birsic E, Fritz LP, Solotky L, Weislogel A, Schlaegel S (2019) Quantitative evaluation of vitrinite reflectance in shale using Raman spectroscopy and multivariate analysis. Fuel 254:115573. https://doi.org/10.1016/j.fuel.2019.05.156

Mählmann RF, Le Bayon R (2016) Vitrinite and vitrinite like solid bitumen reflectance in thermal maturity studies: Correlations from diagenesis to incipient metamorphism in different geodynamic settings. Int J Coal Geol 157:52–73

Mashhadi ZS, Rabbani AR, Kamali MR, Mirshahani M, Khajehzadeh A (2015) Burial and thermal maturity modeling of the middle cretaceous-early miocene petroleum system, Iranian sector of the Persian Gulf. Pet Sci 12(3):367–390

Middleton MF (1982) Tectonic history from vitrinite reflectance. Geophys J Int 68(1):121–132

Mir M, Kamyab M, Lariche MJ, Bemani A, Baghban A (2018) Applying ANFIS-PSO algorithm as a novel accurate approach for prediction of gas density. Pet Sci Technol 36(12):820–826

Mohaghegh SD (2017) Data-driven reservoir modeling. A held course in SPE

Mukhopadhyay PK (1994) Vitrinite reflectance as maturity parameter: petrographic and molecular characterization and its applications to basin modeling. American Chemical Society, Washington

Peters KE (1986) Guidelines for evaluating petroleum source rock using programmed pyrolysis. AAPG Bull 70(3):318–329

Peters KE, Cassa MR (1994) Applied source rock geochemistry: Chapter 5: Part II. Essential elements. In: M 60: The petroleum system—from source to trap, pp 93–120

Peters KE, Hackley PC, Thomas JJ, Pomerantz AE (2018) Suppression of vitrinite reflectance by bitumen generated from liptinite during hydrous pyrolysis of artificial source rock. Org Geochem 125:220–228. https://doi.org/10.1016/j.orggeochem.2018.09.010

Petersen HI, Schovsbo NH, Nielsen AT (2013) Reflectance measurements of zooclasts and solid bitumen in Lower Paleozoic shales, southern Scandinavia: correlation to vitrinite reflectance. Int J Coal Geol 114:1–18

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

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. J Pet Sci Eng 177:236–249

Sabah M, Talebkeikhah M, Wood DA, Khosravanian R, Anemangely M, Younesi A (2019) A machine learning approach to predict drilling rate using petrophysical and mud logging data. Earth Sci Inform 12:319–339

Saremi S, Mirjalili S, Lewis A (2017) Grasshopper optimisation algorithm: theory and application. Adv Eng Softw 105:30–47. https://doi.org/10.1016/j.advengsoft.2017.01.004

Shalaby MR, Jumat N, Lai D, Malik O (2019) 'Integrated TOC prediction and source rock characterization using machine learning, well logs and geochemical analysis: case study from the Jurassic source rocks in Shams Field NW Desert, Egypt. J Pet Sci Eng 176:369–380

Singh A (2017) Application of data mining for quick root-cause identification and automated production diagnostic of gas wells with plunger lift. SPE Prod Oper 32(03):279–293

Snowdon LR (1995) Rock-Eval Tmax suppression: documentation and amelioration. AAPG Bull 79(9):1337–1348

Soleimani B, Bahadori AR, Meng F (2013) Microbiostratigraphy, microfacies and sequence stratigraphy of upper cretaceous and paleogene sediments, Hendijan oilfield, Northwest of Persian Gulf, Iran. Nat Sci 5(11):1165–1182

Staplin FL (1961) Reef-controlled distribution of Devonian microplankton in Alberta. Palaeontology 4(3):392–424

Suggate RP (1959) New Zealand coals, their geological setting and its influence on their properties: New Zealand Geological Survey Bulletin 134. Institute of Geological and Nuclear Sciences, Lower Hutt

Sweeney JJ, Burnham AK (1990) Evaluation of a simple model of vitrinite reflectance based on chemical kinetics (1). AAPG Bull 74(10):1559–1570

Taylor BW (2019) Introduction to management science. Pearson, New York

Taylor GH, Teichmüller M, Davis ACFK, Diessel CFK, Littke R, Robert P (1998) Organic petrology, XVI. ISBN 978-3-443-01036-2

Teichmuller M, Teichmuller R (1982) Stach's textbook of coal petrology. Gebruder Borntraeger, Berlin, Stuttgart, pp 381–413

Vapnik V (2013) The nature of statistical learning theory. Springer, Berlin

Wang H, Wu W, Chen T, Dong X, Wang G (2019) An improved neural network for TOC, S1 and S2 estimation based on conventional well logs. J Pet Sci Eng 176:664–678

Waples DW (1984) Thermal models for oil generation. Adv Pet Geochem 1:7–67

Welte DH, Yalcin MN (1988) Basin modelling: a new comprehensive method in petroleum geology. Organic geochemistry in petroleum exploration. Elsevier, Berlin, pp 141–151

Wilkins RWT, Sherwood N, Li Z (2018) RaMM (Raman maturity method) study of samples used in an interlaboratory exercise on a standard test method for determination of vitrinite reflectance on dispersed organic matter in rocks. Mar Pet Geol 91:236–250

Wilkins RWT, Wang M, Gan H, Li Z (2015) A RaMM study of thermal maturity of dispersed organic matter in marine source rocks. Int J Coal Geol 150:252–264

Wust RAJ, Nassichuk BR, Brezovski R, Hackley PC, Willment N (2013) Vitrinite reflectance versus pyrolysis Tmax data: Assessing thermal maturity in shale plays with special reference to the Duvernay shale play of the Western Canadian Sedimentary Basin Alberta Canada. Soc Pet Eng. https://doi.org/10.2118/167031-MS

Yang XS (2010) A new metaheuristic bat-inspired algorithm. In: González JR et al (eds) Nature inspired cooperative strategies for optimization. Springer, Berlin, pp 65–74

Yang XS (2012) Metaheuristic optimization with applications: demonstration via bat algorithm. In: Proceedings of 5th bioinspired optimization methods and their applications (BIOMA2012)(Eds. B. Filipic and J. Silc), pp. 24–25

Yang XS, Hossein Gandomi A (2012) Bat algorithm: a novel approach for global engineering optimization. Eng Comput 29(5):464–483

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.