International Journal of Innovative Computing, Information and Control
Volume **17**, Number **2**, April **2021**
ICIC International ©2021 ISSN 1349-4198
pp. 539–563

ARTIFICIAL INTELLIGENCE APPROACH TO TOTAL ORGANIC CARBON CONTENT PREDICTION IN SHALE GAS RESERVOIR USING WELL LOGS: A REVIEW

Md Shokor A Rahaman[1,*], Jahedul Islam[1], Junzo Watada[2]
Pandian Vasant[1], Hitmi Khalifa Alhitmi[3]
and Touhid Mohammad Hossain[4]

[1]Department of Fundamental and Applied Sciences
[4]Department of Computer and Information Sciences
Universiti Teknologi Petronas
Seri Iskandar, Perak Darul Ridzuan 32610, Malaysia
*Corresponding author: shokor103072@gmail.com
{ jahedul.islam.cuet; pvasant; touhidmhossain }@gmail.com

[2]IPS Research Center
Waseda University
1-104 Totsukamachi, Shinjuku-ku, Tokyo 169-8050, Japan
junzo.watada@gmail.com

[3]College of Business and Economics
Qatar University
P.O. Box: 2713, Doha, Qatar
Hitmialhitmi@gmail.com

Received July 2020; revised November 2020

Abstract. *The most important element for the exploration and development of oil and oil shale is total organic carbon (TOC). TOC estimation is considered a challenge for geologists since laboratory methods are expensive and time-consuming. Therefore, due to the complex and nonlinear relationship between well logs and TOC, researchers have begun to use artificial intelligence (AI) techniques. Hence, the purpose of this research is to explore new paradigms and methods for AI techniques. First, this article provides a recent overview of selected AI technologies and their applications, including artificial neural networks (ANNs), convolutional neural networks (CNNs), hybrid intelligent systems (HISs), and support vector machines (SVMs) as well as fuzzy logic (FL), particle swarm optimization (PSO). Second, this article explores and discusses the benefits and pitfalls of each type of AI technology. The study found that hybrid intelligence technology was the most successful and independent AI model with the highest probability of inferring properties of oil shale oil and gas fields (such as TOC) from wireline logs. Finally, some possible combinations are proposed that have not yet been investigated.*
Keywords: Artificial intelligence, Pattern recognition, Total organic carbon (TOC), Organic shale, Well logs

1. **Introduction.** In the 1940s, petroleum exploration and production (E&P) first introduced optimization methods. After being introduced, the optimization method has been widely used for estimations and prediction [1]. The optimization methods are mainly categorized into three groups: linear, integer and nonlinear programming techniques.

DOI: 10.24507/ijicic.17.02.539

For the linear constraint and objective function problem, the linear programming method is primarily used. Two examples of linear programming techniques are the interior point and simplex algorithm. Although this method is very popular, it takes many iterations to converge which is a considerable drawback. On the other hand, for problems in which all unknown components are mixed continuous or discrete and integer, the integer programming technique is applicable. The branch and bound method and the cutting plane technique are the two methods by which researchers tackle these problem [2,3]. However, it requires high computational time and cost, which is its main disadvantage. The third approach, i.e., the nonlinear programming technique, is used for optimization problems where the constraint or the objective are nonlinear. This nonlinear programming technique has two main parts: the gradient-free optimization algorithm and the gradient-based optimization algorithm [4]. For searching the steepest descent or ascent direction the gradient based algorithm is used. Again, the gradient-based algorithm searches for the function extremes using numerical or analytical objective functions, including Newton's method, sequential quadratic programming technique, quasi-Newton technique, the steepest descent technique and finite difference techniques [5-7]. As suggested by the name, the computation of the objective function and its constraints is also required in the gradient-based algorithm. However, all objective functions are not differentiable for the following reasons:

- The constraints that are defined by the regions or the objective function are non-differentiable;
- A simulation-based objective function for which the derivative computation requires access to the simulation.

Therefore, the gradient-based optimizer fails because of the lack of a computable derivative. Again, the gradient-based optimizer requires the application of derivative-free techniques. On the other hand, gradient-free optimization or heuristic techniques solve the problems more efficiently by using the domain knowledge. Again, heuristic or gradient-free optimization techniques provide nearly optimal solutions and tend to be fast. However, for finding the genuine value of the optimal solution, these techniques cannot be guaranteed [4,8]. Population-based techniques and trajectory-based techniques are the two methods coming from gradient-free approaches. A population-based technique maintains a population of solutions while a trajectory-based technique takes into account only one solution [8].

The summarization of the introduction section is shown in Figure 1 [9], where several examples of optimization techniques are given.

Efforts to provide a comprehensive interpretation of artificial intelligent concepts have been hampered by volatile ambiguity, but such difficulties have great promise as resources for AI to understand interactions between diverse entities. We do not deny the reality of having. However, this research should be compatible with the concept of AI put forth by McCarthy [10], which describes AI as the science and engineering of designing computer programmes that, when fed into a system, render the system exhibit human intelligence. AI is often referred to as "virtual intelligence", "soft computing", and "computational intelligence" Mohaghegh [11]. According to Rable [12], the observable qualities of AI include reviewing vast data to identify patterns and forecast potential outcomes in the best possible way, thus reducing the duplication of time and energy. As a consequence, there are substantial cuts in running expenses. Based on the available literature, nearly any aspect of the oil and gas discovery and development chain utilizes any or other artificial intelligence technologies in its service.

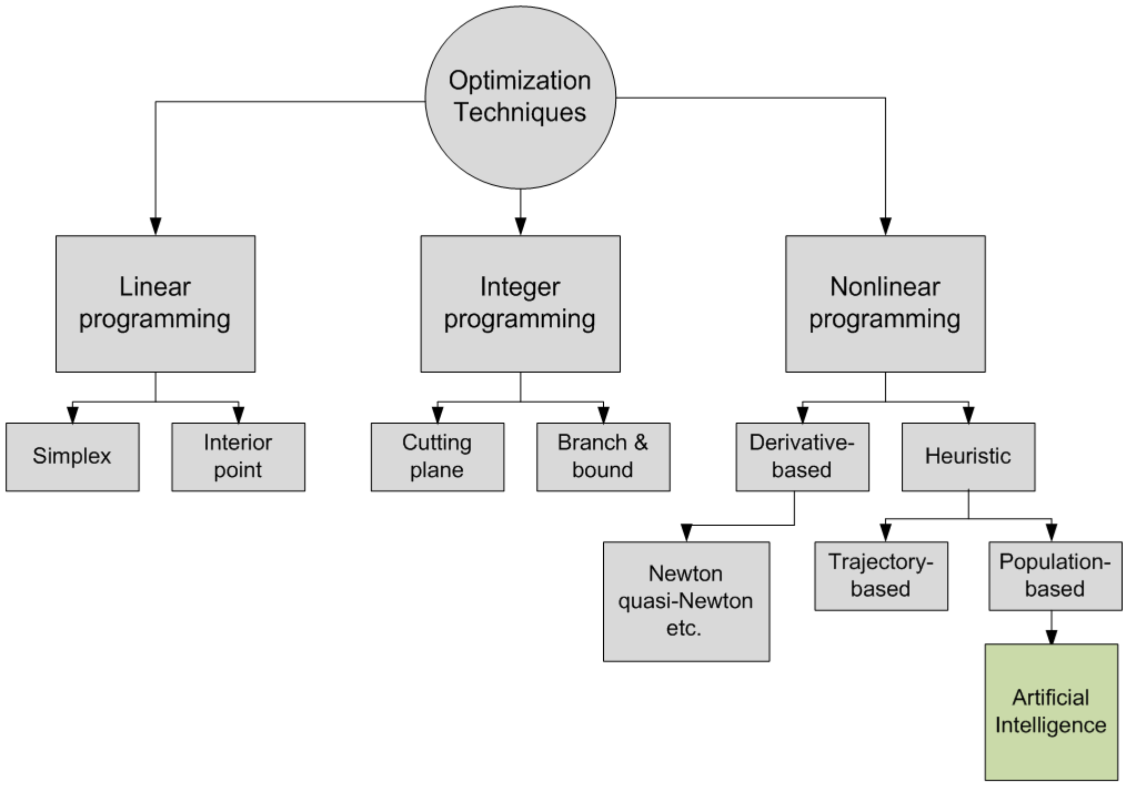

Figure 1. Group of optimization techniques

Since the 1990s, AI techniques have been applied extensively in many engineering and scientific fields including the petroleum industry. Recently, petroleum engineers and geologists have used AI for solving related problems of unconventional hydrocarbon resource evaluation [13,14], bubble point pressure evaluation [15], reservoir characterization [16,17], the optimization of rate of penetration [18], the prediction of real-time change in the rheological parameters of the drilling fluids [19,20], the estimation of rock mechanical parameters [21,22], the optimization of rate of penetration [18], hydrocarbon recovery factor estimation [21,23], the optimization of the drilling hydraulics [24], the prediction of pore pressure and fracture pressure [25,26], the evaluation of the wellbore casing integrity [27,28] and others.

The shale reservoir characterization process requires such very high degree of prediction accuracy that any deviation from expectation may result in huge losses and wasted efforts through enormous man-hours and huge investments. Conversely, a little improvement in the prediction accuracies will have multiplicative effect on current exploration and production activities. Present prediction accuracies have remained acceptable in the oil & gas industry, but there is always the quest for better and more reliable results. In view of this, there is the need for the hybridization of those techniques with traits that are strong enough to be used to complement the performance of other techniques for increased performance in terms of higher prediction accuracies, reduced prediction errors, and faster execution. However, no comprehensive study of the application of different AI techniques to estimate TOC content from wireline/well logs in shale gas reservoirs has been conducted. Therefore, we offer this summary of the most pertinent literature on artificial intelligence (AI).

2. **Problem Formulation.** Unconventional resources (i.e., low permeability-porosity reservoir) such as tight oil and shale gas extended their recovery ability of hydrocarbon

for recent advancement in multistage fracturing and horizontal drilling. Among the shale reservoir properties total organic carbon (TOC) is one of the most important parameters that directly affect hydraulic fracture design as well as reserve estimation [29-31]. Again, a shale formation geomechanical property considerably has been affected by organic matter which are like all other rock matrix components. For developing such a reservoir these components are important [32]. Further, a reservoir organic porosity is controlled by carbon content and maturity. Hence, the maturity and carbon content affect the organic matter by which gas is being absorbed [33,34].

TOC present in shale reservoir also controls permeability, texture, reservoir wettability, and microstructure [31,32]. Thus, for characterizing the organic matter present in a shale formation, a reliable method is much needed for hydrocarbon exploration and production [35,36].

Currently, the application of empirical correlation estimates TOC by using well logs or wireline logs data. However, for certain conditions and formation, these empirical correlations are developed. Again, these correlations are based on specific assumptions.

The first empirical equation for TOC is developed by the Schmoker [37] where the Devonian formation data is used. For determining the *vol*% of TOC, Equation (1) can be used. Again, by the explanation of Schmoker [37], *wt*% of TOC can be retrieved from *vol*% of TOC value.

$$TOC(vol\%) = \frac{(\rho_B - \rho)}{1.378} \tag{1}$$

where ρ and ρ_B respectively denote the rock bulk density in g/cm^3 and the rock density is without including the organic matter in g/cm^3.

The above model had been modified by Schmoker [37] for being used in the Bakken shale formation where the same pyrite-organic matter relationship was assumed, Equation (2):

$$TOC(wt\%) = \frac{[(100\rho_0)(\rho - 0.9922\rho_{mi} - 0.039)]}{[(R_\rho)(\rho_0 - 1.135\rho_{mi} - 0.675)]} \tag{2}$$

where ρ_{mi} and ρ_0 are the volume-weighted average density of the pore fluid and grain in g/cm^3 and the organic matter density in g/cm^3 respectively. Again, the weight percent ration of the organic matter to organic carbon is R.

For TOC, a more generalized model for the Bakken formation is developed by Schmoker and Hester [38] which had been represented by a basic equation (Equation (3)). In this equation $\rho_0 = 1.01$ g/cm^3, $R = 1.3$ and $\rho_{mi} = 2.68$ g/cm^3 are assumed.

$$TOC(wt\%) = \frac{77.44}{\rho} - 28.7 \tag{3}$$

Passey et al. [39] developed a simplified method for the estimation of TOC (Equations (4) and (5)), which found wide application in the field of unconventional resource evaluation.

$$\Delta \log R = \log_{10}\left(\frac{R}{R_{baseline}}\right) + 0.02 \times (\Delta t - \Delta t_{baseline}) \tag{4}$$

$$TOC = \Delta \log R \times 10^{(2.297 - 0.1688 \times LOM)} \tag{5}$$

where the level of maturity is LOM. R, $\Delta \log R$ and $R_{baseline}$ are the resistivity, the logs separation and the resistivity of the base formation in ohm m respectively, Δt and $\Delta t_{baseline}$ are the sonic transient time and the basic sonic transient time in µs/ft.

However, for the other formation, these empirical equations failed to evaluate the TOC content because these empirical equations only can work for those formations for which they are made for [14,36]. With continuous improvements, many milestones have been achieved by the researchers for estimating TOC content. For TOC content evaluation, the direct laboratory method is still important though this process can only obtain limited TOC data. Again, the source rock maturity and the background value of TOC content were different for the various researcher's area which impacted significantly on the prediction for evaluation of the TOC content. Again, a complicated nonlinear relationship is seen between the logging data and TOC content. Due to the complicated nonlinear function relation between the logging information and TOC content, approximation of the real function relationship by simple linear regression is difficult and an alternative approach is required to estimate TOC content from well logs. The available empirical correlations developed based on the linear regression were made to learn to estimate the TOC in a particular Formation. Therefore, to apply the same correlation in a different formation, the correlation must be modified according to the properties of the target formation. An important aspect observed in the recent studies is that for the nonlinear implicit function, artificial intelligence proves to be more prominent. For this reason, many artificial intelligence techniques have been proposed for the TOC content estimation that has become a useful tool in shale oil and gas exploration. In fact, the use of robust artificial intelligence methods approaches has been introduced and successfully employed in many petroleum engineering fields, such as unconventional hydrocarbon resources evaluation [13,14], bubble point pressure evaluation [15], reservoir characterization [16,17], optimization of rate of penetration [18], prediction of real-time change in the rheological parameters of the drilling fluids [19,20], estimation of rock mechanical parameters [21,22], optimization of rate of penetration [18], hydrocarbon recovery factor estimation [21,23], optimization of the drilling hydraulics [24], prediction of pore pressure and fracture pressure [25,26], evaluation of the wellbore casing integrity [27,28]. These methods combine the accuracy of numerical models with the simplicity of analytical approaches, while it is free from constraints of a certain function form. So, by the current research results, AI strategies have worked for predicting the TOC content.

From Table 1 it can be seen that researchers have used five countries as a study area for shale gas reservoir characterization where China is used mostly as a country and Barnett Shale from the USA is used most individually by the researchers as a study area. Again, the most recent techniques for estimating TOC content from multiple wireline logs is summarized in Table 1 which are artificial neural network (ANN), support vector machines (SVMs), hybrid intelligent systems (HISs), fuzzy logic (FL), particle swarm optimization (PSO) algorithm, convolutional neural network (CNN) and extreme learning machine (ELM). The purpose of this table is to give an idea to the readers about the recent research trends of utilizing various AI techniques. However, both ANN and SVM have some drawbacks during the training process because it easily stuck in the local optimum and suffers overfitting, thus substantially reducing the prediction accuracy. Again, in SVM there are still some parameters need to be optimized. Further, extreme learning machine, fuzzy theory, sequential Gaussian simulation (SGS) are mainly based on statistical theory. Furthermore, some AI techniques had come up with some certain limitations and challenges that would not make its application desirable in certain conditions such as small, sparse, limited, and missing data scenarios [40-42], and model complexity and high data dimensionality conditions [43,44]. The "no free lunch" theory [45] also holds true as no single one of the AI techniques could be considered as being the best to solve all problems in all data and computing conditions. Since each of the techniques has its limitations and challenges associated with its strengths, there have been few research attempts in the area of hybrid intelligent systems (HISs) [41,46,47] to have better generalization than individual AI techniques. Hence, this study attempts to provide a comprehensive review of AI techniques on TOC estimation from wireline logs.

Table 1. Artificial intelligent techniques used to predict reservoir properties from shale reservoir

No	Authors	Study area	Application	The technique(s)
1	[13]	Barnett Shale, USA	Prediction of TOC content	ANN, SaDE-ANN
2	[48]	Bohay Bay Basin, China	Prediction of TOC, S1 and S2	ANN-BP, CNN
3	[49]	Barnett Shale, USA	Prediction of TOC content	TSK-FIS, M-FIS, SVM, FNN
4	[50]	Shams Field, NW Desert, Egypt	Prediction of TOC content	ANN
5	[51]	Beibu Gulf Basin, China	Prediction of TOC content	SVM, PSO-SVM, MLP-NN
6	[52]	Canning Basin, Australia	Prediction of TOC, S1, S2 and S3	ANN
7	[53]	Ordos Basin, China	Prediction of TOC content	SAGA-FCM, BPNN, LSSVM, PSO-LSSVM
8	[54]	Tonghua Basin, China	Prediction of TOC content	LSSVM, ANN-BP, PSO-LSSVM
9	[55]	Canning Basin, Australia	Prediction of TOC, S1 and S2	ANN
10	[56]	Zagros Fold-Thrust Belt, Iran	Prediction of TOC content	ANN
11	[57]	–	Prediction of TOC content	IHNN, BP-ANN, BP-Adaboost, KELM, SVM
12	[58]	–	Prediction of TOC and FI content	GA, MLR, HML
13	[14]	Barnett Shale, USA	Prediction of TOC content	ANN
14	[59]	Persian Gulf Basin, Iran	Prediction of TOC content	Fuzzy logic, K-means clustering, ANN, SVM
15	[60]	Sichuan Basin, China	Prediction of TOC content	MLP-ANN, ELM
16	[61]	–	Prediction of TOC content	SVM
17	[62]	Barnett Shale, USA	Prediction of TOC content	FL, MLP-NN
18	[63]	Barnett Shale, USA	Prediction of TOC content	MLP-ANN
19	[64]	South Pars Gas Field, Iran	Prediction of TOC content	ANN

The efficiency of AI techniques on evaluating the TOC was evaluated based on the statistical indicators, i.e., mean absolute error (MAE, Equation (6)), mean square error (MSE, Equation (7)), coefficient of determination (R^2, Equation (8)), absolute average deviation (AAD, Equation (9)), root mean square error (RMSE, Equation (10)) and mean absolute percentage error (MAPE, Equation (11)). These statistical indicators mathematical equations have been described below:

$$MAE = \frac{1}{N \times p} \sum_{i=1}^{p} \sum_{j=1}^{N} |T_{ij} - L_{ij}| \tag{6}$$

$$MSE = \frac{1}{N \times p} \sum_{i=1}^{p} \sum_{j=1}^{N} (T_{ij} - L_{ij})^2 \tag{7}$$

$$R^2 = \frac{\sum_{i=1}^{n} (Y_{i,m} - Y_{i,e})^2}{\sum_{i-1}^{n} (Y_{i,m} - \bar{Y}_{i,m})^2} \tag{8}$$

$$AAD = \frac{1}{N} \sum_{i=1}^{N} \left| T_{ij} - \overline{T_i} \right| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N \times p} \sum_{i=1}^{p} \sum_{j=1}^{N} (T_{ij} - L_{ij})^2} \tag{10}$$

$$MAPE = 100 \times \frac{1}{N \times p} \sum_{i=1}^{p} \sum_{j=1}^{N} \left| \frac{T_{ij} - L_{ij}}{T_{ij}} \right| \tag{11}$$

where p, N represent the number of data set patterns and the number of output units. Again, T_{ij} are the target values and L_{ij} are the output values. Furthermore, $\bar{Y}_{i,m}$, $Y_{i,e}$, $Y_{i,m}$ and n represent average laboratory measured values, well logging parameters, laboratory measured values and number of samples respectively. The model will perform better if the value of MSE, AAD, RMSE, MAPE and MAE is low. Conversely, higher value of coefficient of determination (R^2) means its value is closer to 1 which makes the regression line fits the data well and better model performance.

3. **Artificial Intelligent Techniques for TOC Estimation.**

The concept of AI can be described as the use of software that is ultimately designed to exhibit superintelligence by recognizing patterns from a given set of data and information and thus be able to draw an inference that could be used in solving real-world problems like reservoir characterization.

AI algorithms are classified as intelligent because they are supposed to be to recognize patterns in models and dataset(s), learn those patterns, and proffer solutions to problems based on the patterns they have recognized.

The application of AI is made possible by the availability of large volumes of data (big data), and the accessibility of a huge amount of data points from multiple logs makes the application of AI in shale gas reservoir characterization more promising. In this section, a survey of state-of-the-art researches involves the application of AI techniques in shale gas reservoir characterization. Firstly, the fundamental of these artificial intelligence techniques are highlighted; secondly, their application in TOC content estimation in shale gas reservoir are discussed.

3.1. **Artificial neural network (ANN).**

ANN is a supervised training intelligent system for solving nonlinear problems, which is developed under the category of artificial intelligence (AI) to provide a brain-like tool. An input layer, a hidden layer, and an

TABLE 2. Summary of researchers on TOC content prediction where ANN was applied

Authors	Type of study conducted	ANN method	ANN architecture	Study field	Input parameters	Output parameters	Errors (Performance evaluation criteria)				
							RMSE	**AAPE**	$\mathbf{R^2}$	**MAE**	**MSE**
[48]	TOC, S1 and S2 prediction	BP-ANN	4-10-3	Bohay Bay Basin, China	DEN, RT, CNL, AC, GR	TOC, S1, S2	–	–	0.75 (for TOC)	–	–
[13]	TOC prediction	ANN	–	Barnett Shale, USA	GR, DT, RD, RHOB	TOC	–	19	0.90	–	0.072
[49]	TOC prediction	FNN	–	Barnett Shale, USA	GR, DR, DT, RHOB	TOC	–	12.02	0.87	–	–
[50]	TOC prediction	ANN	–	Shams Field, NW Desert, Egypt	LLD, RHOB, GR, NPHI	TOC	–	–	0.42	–	–
[55]	TOC, S1, S2 prediction	ANN	4-4-3	Canning Basin, Australia	RD, RHOB, GR, SP	TOC, S1, S2	–	–	0.79 (for TOC)	–	0.16631 (for TOC)
[53]	TOC prediction	BP-ANN	9-8-1	Ordos Basin, China	SP, GR, DTC, RT, U, KTH, TH, DEN, CNL	TOC	2.059	–	0.88	–	–
[54]	TOC prediction	BP-ANN	–	Tonghua Basin, China	SP, CNL, DTC, RT, U, GR, DEN, K, TH	TOC	0.093	–	0.90	–	–
[57]	TOC prediction	IHNN	–	–	RD, TH, KTH, CNL, PE, GR, AC, U, DEN	TOC	–	–	–	0.303	–
[52]	TOC, S1, S2 and S3 content prediction	MLP-NN	11-10-4	Canning Basin, Australia	GR, DT, NPHI, LLD, LLS, MSFL, SP, CALI, FCINL, RHOB	TOC, S1, S2, S3	–	–	0.95 (for TOC)	–	–
[56]	TOC prediction	BP-ANN	2-5-1	Zagros Fold-Thrust Belt, Iran	LLD, DT	TOC	–	–	0.89	–	–
[59]	TOC prediction	ANN	9-35-1	Persian Gulf Basin, Iran	NPHI, DT, LLD, LLS, GR, RHOB, URAN, POTA, THOR	TOC	–	–	0.90	–	–
[14]	TOC prediction	MLP-NN	4-5-1	Barnett Shale, USA	GR, ILD, Δt, ρ	TOC	–	–	0.89	–	0.0135
[63]	TOC prediction	MLP-NN	5-9-1	Barnett Shale, USA	GR, NPOR, RHOB, DTC, DSTM	TOC	–	–	–	–	–
[64]	TOC prediction	BP-ANN	4-9-1	South Pars Gas Field, Iran	Rlld, Δt, FDC, NPHI	TOC	–	–	–	–	0.02

output layer consist in ANN model. Among those layers, each layer is interconnected with many neurons with a specific function, such as sigmoid, purelin. A full connection is established between each node of the next layer. It can not only handle the difficult non-linear problems in engineering research but also have a good effect on data classification and prediction. For the fundamental details about ANNs, a good read is [65].

Cases where ANN was applied in TOC prediction. Researchers who utilized artificial neural network (ANN) for TOC content estimation have been summarized in Table 2. The research paper between 2011 to 2019 is available in this survey paper because tracing the earliest research paper was difficult. The areas lie in shale reservoir characterization where ANN has been applied for predicting TOC content from wireline logs. In this table seven things are highlighted which are respectively, author name, the type of study conducted by each researcher, the ANN method used, the ANN architecture, the input parameters, and the output parameters and lastly the magnitude of statistical indicators (performance evolution criteria), e.g., RMSE, R^2, RMAE, MSE, AAPE, are given for evaluating the result of the developed ANN model. Followings are given from the summery:

- Most of the researchers (about 78% from Table 2) used ANN for multiple inputs with respect to single output except these three authors where [48,55] carried their works by having 3 output which is TOC, S1 and S2 respectively and [52] carried his works by having 4 output including S3 content. The correlation coefficient (R^2) is the most widely used performance evaluation criterion. From Figure 2, the average value of the correlation coefficient (R^2) was approximately 85% (across the research works reviewed) which makes ANN as a good model.
- The aspects of shale reservoir properties estimation show that ANN was mostly applied for prediction of TOC content. These accounted for 74% of the researchers.
- Researchers mainly used five-country data (namely Australia, China, Egypt, Iran and USA) as a study area for evaluating reservoir properties (for predicting TOC content from shale reservoir).

3.2. **Fuzzy logic.** Fuzzy logic (FL) was first proposed by Zadeh [66] as a generalization of the set theory of two- and many-valued logic. It deals with systems having

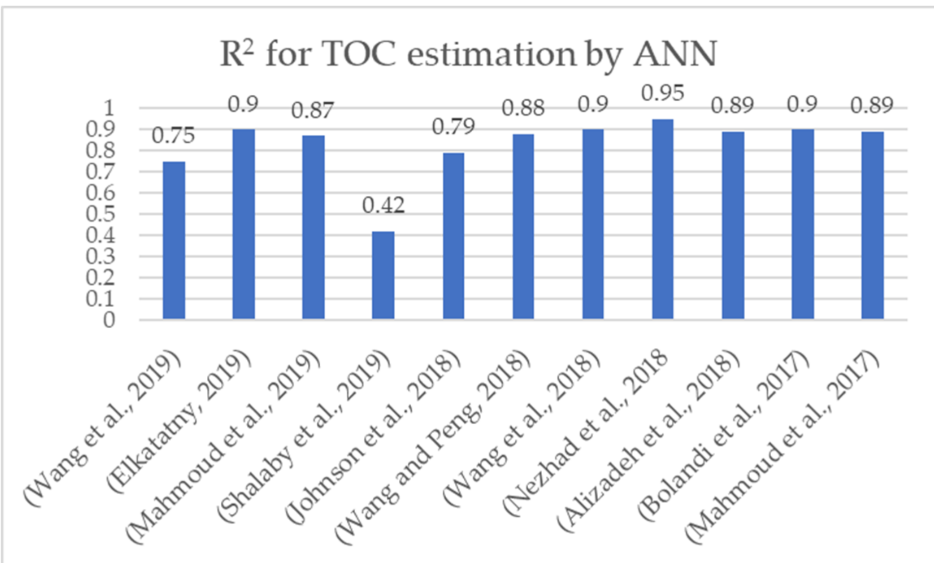

FIGURE 2. R^2 for TOC estimation by ANN

non-crisp boundaries; these systems exhibit characteristics such as being hazy and unclear/ambiguous [67]. Suppose x is a member of the total space X. In the crisp logic, the characteristic function x_A defines the set A on total space (X) by mapping the total space to the set $\{0, 1\}$, as given by the following expression [67]:

$$x_A : X \to \{0, 1\} \tag{12}$$

$$x \to x_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases} \tag{13}$$

Therefore, the characteristic function can take the value of 1 if x belongs to A, while the function is 0 if it is not a part of A. In the FL, Equation (9) is equivalent to the following relationship:

$$m_A : X \to [0, 1] \tag{14}$$

$$x \to m_A(x) \tag{15}$$

in which, m_A represents the membership function [66]. The membership function in the literature is shown with different notations such as fA (in the original work conducted by Zadeh) and μA [66]. The difference between the characteristic function in Equations (12) and (13) and the membership function in Equations (14) and (15) is that the mapping of space X is a Boolean set ("0" and "1", or "yes" and "no") in the crisp logic and a domain ($0 \leq m_A \leq 1$) in the FL. In the work carried out by Zadeh [66], a "class" is represented by a continuum of "grades-of-membership". [68] is a very good read for fundamental and application details of FL.

Cases where fuzzy logic was applied in TOC prediction. Research papers reviewed on the use of fuzzy logic in TOC content estimation from well logs in shale reservoir characterization are seen in Table 3. From the table it can be seen that FL had a very good accuary to predict TOC content from wireline logs where the maximum and minimum accuracy were 99.2% and 91.8% respectively. Again, FL had been used to predict TOC content in two countries (Iran and USA) study area.

TABLE 3. Summary of researchers on TOC content prediction where fuzzy logic (FL) was applied

Authors	Type of study conducted	Study area	Input parameters	Output parameters	Errors (Performance evaluation criteria)		
					AAPE	$\mathbf{R^2}$	**MSE**
[49]	TOC prediction	Barnett Shale, USA	DR, DT, GR, RHOB	TOC	11.20% (for TSK-FIS) 11.10% (for M-FIS)	0.918 (for TSK-FIS) 0.933 (for M-FIS)	–
[59]	TOC prediction	Persian Gulf Basin, Iran	DT, GR, NPHI, RHOB, LLS, LLD	TOC	–	0.9425	–
[62]	TOC prediction	Barnett Shale, USA	GR, RHOB, NPOR, DTCO, DTSM	TOC	–	–	–
[64]	TOC prediction	South Pars Gas Field, Iran	Rlld, Δt, FDC, NPHI	TOC	–	0.992	0.001

3.3. **Particle swarm optimization.** Particle swarm optimization is an optimization method that is based on swarm intelligence. From the predation behavior of birds, the idea of PSO is derived. Kennedy and Eberhart first developed the particle swarm optimization in 1995 [69].

Initially, PSO starts with a population of random solution where each solution is called particle. For searching into the D-dimensional solution space, velocity is assigned to each solution. For determining each particle current position, the fitness function is used. Again, after finding the best position, the particles record it. The equation for getting the value of position and velocity of all the particles is given below:

$$v_{i,j}^{k+1} = \omega \cdot v_{i,j}^{k} + c_1 r_1 \left(xpbest_{i,j}^{k} - x_{i,j}^{k} \right) + c_2 r_2 \left(xgbest_{i,j}^{k} - x_{i,j}^{k} \right) \tag{16}$$

and

$$x_{i,j}^{k+1} = x_{i,j}^{k} + v_{i,j}^{k+1} \tag{17}$$

where $v_{i,j}^{k}$ is the velocity vector and $x_{i,j}^{k}$ is the position of jth component of ith particles. Again, the two random numbers r_1 and r_2 in the kth iteration with a range between $(0, 1)$ increase the search randomness. Further, the global and local learning rates are controlled by the non-negative acceleration factors c_1 and c_2. The inertia weight ω which is also a non-negative is considered as a constant value which range is between 0 and 1 for adjusting the searching range. $xpbest_i$ is the ith particle's best position and $xgbest$ is all particles' best position. From Equation (16), the particle velocity gets updated. Equation (16) can be divided into three parts which are respectively "inertia part", "self-cognition" and "social experience". The first part which is "inertia part" represents the memory of the previous particles' velocity. The following part which is "self-cognition" can be recognized as the current position of ith particle and its best position and the final part which is "social experience" is the distance between the group's current position of ith particle and best position. The parameters ω, c_1, c_2 are the part of the mentioned three parts where they contribute to updating the velocity.

Cases when PSO was used to predict TOC content in shale gas reservoir. Table 4 highlights the research activities of various researchers where they used PSO optimization technique in conjunction with the support vector machine to improve the prediction of TOC content from well logs in shale reservoir characterization. In this work PSO was basically hybridized with SVM for increasing the efficiency of SVM to predict TOC content. Here, the maximum accuracy of PSO-SVM was 91% and minimum was 89%.

Table 4. Summary of researchers on TOC content prediction where PSO was applied

Authors	Type of study conducted	Type of HYBRID	Study area	Input parameters	Output parameters	Errors (Performance evaluation criteria)		
						R²	**RMSE**	**MAE**
[51]	TOC prediction	PSO-SVR	Beibu Gulf Basin, China	DT, GR, M2RX, SP, ZDEN	TOC	0.89	0.582	0.477
[53]	TOC prediction	PSO-LSSVM	Ordos Basin, China	SP, GR, TDC, RT, U, KTH, TH, DEN, CNL	TOC	0.95	0.3383	–
[54]	TOC prediction	PSO-LSSVM	Tonghua Basin, China	SP, CNL, DTC, RT, U, GR, DEN, K, TH	TOC	0.91	0.0841	–

3.4. **Support vector machines (SVMs).** Support vector machines (SVMs) were first introduced in 1960s [70]. Later in 1990s, the researcher started using the idea of SMV more often when they saw the superior performance over ANN in different classification problems which makes SVM successfully accepted technique [71]. SVM goes under the machine learning category of nonparametric models (supervised learning). The potentiality of SVMs in regression, clustering, classification, and forecasting applications is excellent [72]. Again, for high dimensional data or noisy data, SVM performance is good [73]. Further, SVM has been utilized to solve classification, regression, prediction and function approximation, among others. A least square support vector machine (LSSVM) was developed by [74] for solving the linear equations. Unlike the standard SVM, LSSVM considers equality-type constraints. LSSVM is relatively easy to train and its modeling performance is good.

Cases where support vector machine was applied in TOC prediction. Table 5 highlights the research endeavors of various researchers on the application SVMs for estimating TOC content from well logs in shale reservoir characterization. Practical problems tackled here with the use of SVMs for TOC content prediction from wireline-logs. In the TOC content estimation, the maximum accuary of SVM was 93.7% while the minimum was 77.92%. Again, from the table it can be seen that the Gaussian function is the most common kernel function that has been used by most researchers in SVM.

TABLE 5. Summary of researches on TOC prediction to which support vector machine (SVM) was applied

Authors	Type of study conducted	SVM kernel function used	Input parameters	Output Parameters	Errors (Performance evaluation criteria)		
					R^2	RMSE	MAE
[51]	TOC prediction	Linear, Polynomial, Gaussian	DT, GR, M2R2, M2RX, SP, ZDEN	TOC	0.7792 (Gaussian) 0.696 (Linear) 0.713 (Polynomial)	0.607 (Gaussian) 1.119 (Linear) 0.999 (Polynomial)	0.545 (Gaussian) 0.887 (Linear) 0.825 (Polynomial)
[49]	TOC prediction	Gaussian	DR, DT, GR, RHOB	TOC	0.867	–	–
[54]	TOC prediction	Least square	RT, DTC, GR, DEN, CNL, SP, KTH, TH, U	TOC	0.926	0.078	–
[59]	TOC prediction	Radial Basis Function	GR, DT, LLD, LLS, NPHI, RHOB, POTA, THOR, URAN	TOC	0.937	–	–

3.5. **Convolutional neural network.** The convolutional neural network is an improvement of conventional ANN. Again, in the field of deep learning CNN is an end-to-end learning model. The difference between CNN and ANN is that CNN interfaces the nearby two layers by means of local connection, global sliding, and weight sharing while in ANN all neurons in each layer are completely associated with every neuron in the following layer. CNN network structure is increasingly versatile and more straightforward. Complete CNN models contain an input layer, output layer, convolutional layers, and pooling layers where the convolutional and pooling layer are the most significant. Figure 3 [48] is showing the schematic structure of CNN.

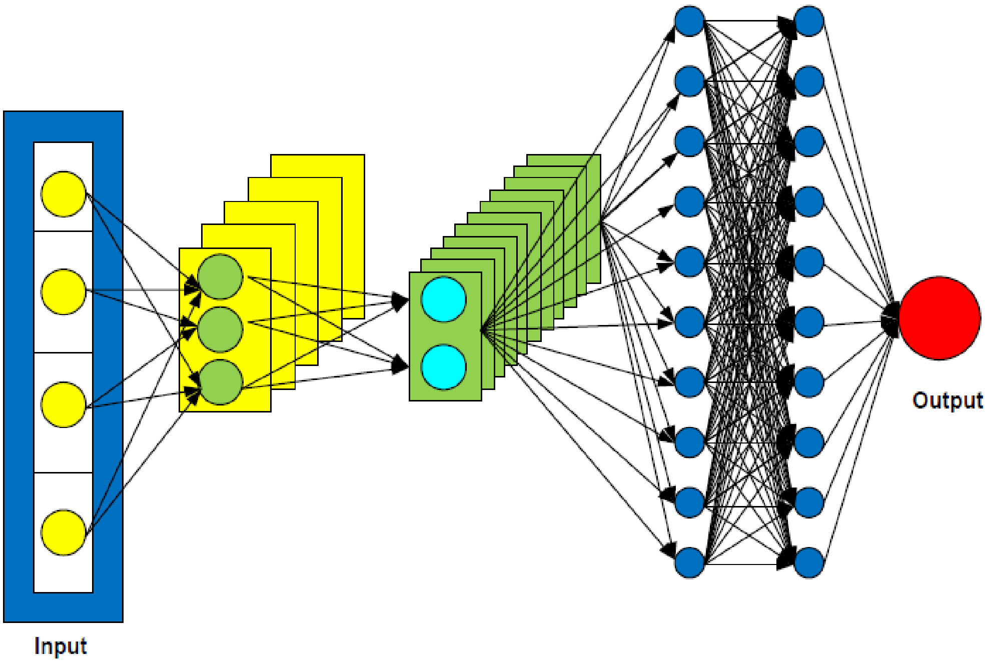

FIGURE 3. Schematic diagram of a convolutional neural network (CNN) structure

Cases where CNN was applied in TOC content prediction. As at the time of this review, the only recorded case of using CNN in TOC content prediction was the work done by [48] where they used the CNN method to improve the prediction of the TOC, S1 and S2 content in shale gas reservoir from wireline logs. Again, CNN is used for removing the limitation of BP-ANN and $\Delta \log R$ method. Further, CNN identified the favorable oil generation layers of Shehejie Formation.

3.6. **Hybrid intelligent systems.** A combination of two or more AI techniques makes a hybrid intelligent system which cooperatively works together for better performance by forming a single functional entity [75,76]. For overcoming the weaknesses of AI techniques this system combines the strengths of multiple AI techniques. This system is increasing their popularity for getting success in many real-world complex problems.

By combining two or more techniques which form a single overall technique is called a hybrid intelligent system [77,78]. In this system, a combination of algorithms such as soft computing methodologies, data mining, and different theoretical backgrounds occurs. Hence, the individual performance of AI techniques is being boosted by the hybridization of AI techniques which achieves more success in dealing with complex problems. There are different flavours in hybrid intelligent systems such as cooperative architecture, feature selection, and optimization.

In Figure 4, HIS modelling framework is shown. This figure shows how each technique gives its contribution to its respective part for making a single overall technique. By combining cooperative effort and synergetic strength each technique increases their strength for solving a problem that suppresses the weakness of the respective techniques.

3.6.1. *Cases where hybrid system was applied in TOC prediction.* In Table 6, researchers who have used a hybrid intelligent system for TOC content prediction in shale reservoir are highlighted. For instance, since DE and PSO are optimization algorithms, hybrids of PSO and SVM to form PSO-LSSVM as well as DE and ANN to form SaDE-ANN would

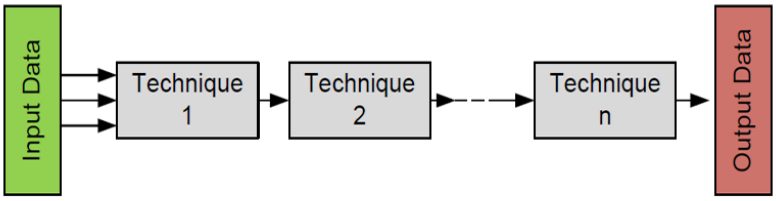

FIGURE 4. Framework of hybrid intelligent system

TABLE 6. Summary of researches on TOC prediction to which hybrid intelligent systems were applied

Authors	Type of study conducted	Type of HYBRID	Number of data points	Input parameters	Output parameters	Errors (Performance evaluation criteria)				
						R^2	**RMSE**	**MAE**	**RMAE**	**MSE**
[51]	TOC prediction	PSO-SVR	–	DT, GR, M2RX, SP, ZDEN	TOC	0.890	0.582	0.477	–	–
[13]	TOC prediction	SaDE-ANN	489	DT, GR, RHOB, RD	TOC	0.98	–	–	–	–
[53]	TOC prediction	PSO-LSSVM	–	SP, GR, TDC, RT, U, KTH, TH, DEN, CNL	TOC	0.9451	0.3383	–	–	–
[54]	TOC prediction	PSO-LSSVM	215	SP, CNL, DTC, RT, U, GR, DEN, K, TH	TOC	0.911	0.0841	–	–	–
[57]	TOC prediction	IHNN	132	RD, TH, KTH, CNL, PE, GR, AC, U, DEN	TOC	–	–	–	0.303	0.294

essentially make their performance better off than when used singly. Below, PSO-LSSVM and SaDE-ANN have been discussed briefly.

3.6.2. *PSO-LSSVM.* Particle swarm optimization (PSO) comes under swarm intelligence which itself is a global optimization algorithm. From the study of the predation behavior of birds, the idea of PSO is derived which is developed by Kennedy and Eberhart [69]. The indirect communication between the individuals makes the optimal solution. The simulation is being foraging the process of bird flocks with this method.

In order to improve the learning ability and generalization ability of LSSVM, this study used a PSO algorithm to realize the global optimization of the LSSVM parameters, and the optimization process is shown in Figure 5 [54].

3.6.3. *SaDE-ANN.* SaDE-ANN is a hybrid model of differential evolution (DE) and artificial neural network (ANN). For improving the parameters (for example learning rate, momentum, number of neurons, number of layers) of ANN, the differential evolution is hybridized with ANN which makes it self-adaptive differential evolution-artificial neural network (SaDE-ANN). This study used a self-adaptive differential evolution (SaDE) to optimize the ANN model and the process is shown in Figure 6 [79].

4. Critical Analysis.

4.1. **Strength and limitation of artificial intelligence techniques.** In this survey, seven AI techniques have been highlighted. It is relevant to ask for all conditions or circumstances if any of them can be said to be 100% suitable and flawless. Wolpert and Macready [45] had propounded a theorem which is "No Free Lunch Theorem for

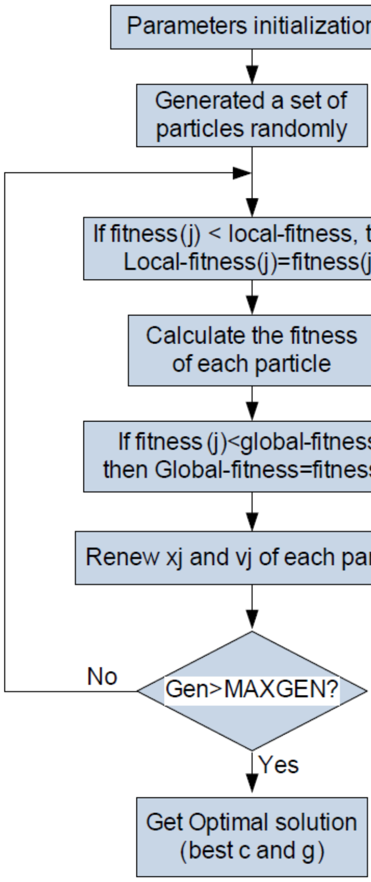

Figure 5. Flow chart of PSO-LSSVM

Optimization (NFLTO). Anifowose et al. [80] opine in support of the NFLTO that there is not a single all-encompassing AI approach that will viably address all difficulties in all data. In Table 7, specifically four methods which are ANN, SVM, fuzzy logic, and PSO were benchmarked on six criteria. These six criteria are robustness against noise, susceptibility to overfitting, ability to self-organize, convergence speed, generalization ability, and data volume requirements. However, from Table 7, it is seen that ANN, SVM, fuzzy logic, and PSO are robust against noise, while fuzzy logic has a better speed of convergence compared with the rest three algorithms. Both ANN and SVM can generalize while ANN requires huge data to predict complex phenomena and can self-organize while SVM requires small data volumes.

From Table 8, it is seen that the researchers for the most part looked at ANN against SVM. With the performance criteria of better execution with lower RMSE, MAPE, MSE and MAE qualities, and high R^2 values, it is seen that SVM tends to perform better in most of the cases referred to. Further, if there should be an occurrence of examination between hybrid intelligent systems with ANN, SVM, fuzzy logic, hybrid performed superior to the ANN, SVM, fuzzy logic, and so on utilized alone.

4.2. **Methodology of TOC estimation using artificial intelligence.** There is an extensive application of artificial intelligence (AI)-based solution to complex engineering problems. In this section, we focus on the problems related to TOC content prediction in shale reservoir.

- Wang et al. [48] proposed convolutional neural network (CNN) for the first time for TOC, S1 and S2 estimation from Dongying Depression, Bohai Bay, China. In their

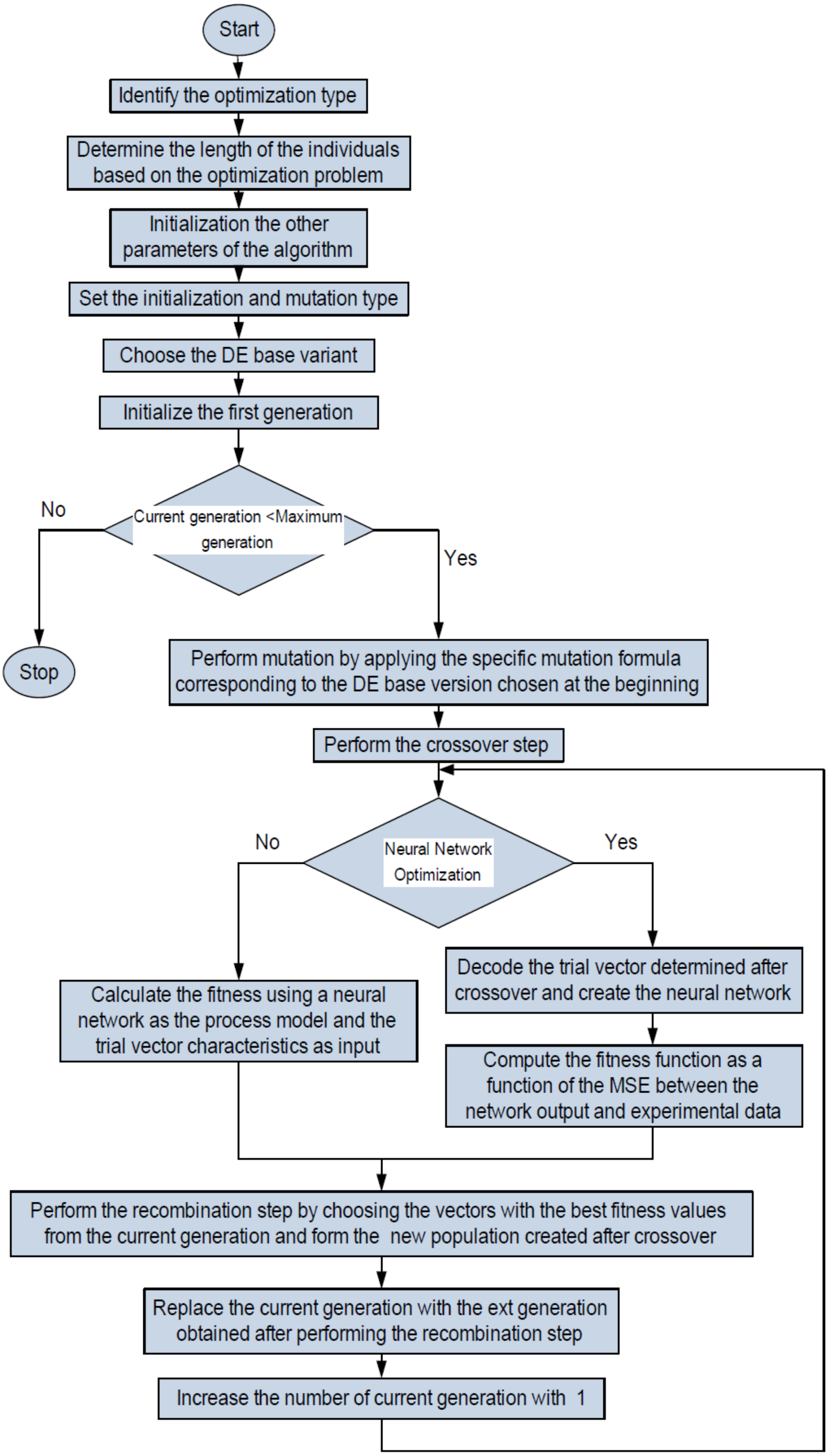

FIGURE 6. Flowchart of SaDE-ANN

Table 7. Summary of the strengths and weaknesses of various AI methods

Benchmark	ANN	FUZZY	SVM	PSO
Robustness against noise	High	High	High	High
Prone to overfitting	Yes, but depends on how the training is done	–	No	–
Self-organization	Yes	–	–	No
Speed of convergence	Slow	Fast	–	High
Ability to generalize	Yes	–	Yes	–
Data requirements	Huge data required	–	Small data required	–

work, the prediction accuracy of CNN was higher than that of BP-ANN and $\Delta \log R$ by removing their limitation.

- A self-adaptive differential evolution (SaDE) optimizing algorithm has been used by the author Elkatatny [13] for finding the best combination of ANN parameters which is the first goal of their work. The result of SaDE-ANN was promising with high accuracy (0.99 correlation coefficient and 6% AAPE).
- In Shalaby et al.'s [50] work, the mathematical model outperformed the machine learning models (ANN and Random Forest) with a R^2 value of 0.9 and 0.4 respectively.
- Four AI techniques (TSK-FIS, M-FIS, FNN and SVM) were developed by Mahmoud et al. [49]. From the result, it was observed that FNN outperformed other techniques for predicting TOC content with 12.02% AAPE and 0.879 correlation coefficient (R).
- Rui et al. [51] had used a support vector machine for continuous TOC content estimation from well logs. Then they used PSO for optimizing SVM. From their result, it is observed that PSO-LSSVM was still better than the other AI models.
- Nezhad et al. [52] used two methods based on machine learning and geostatic tools. From the SGS diagrams, ANN has shown the superiority of producing similar results to the raw data.
- Wang and Peng [53] optimized the data by using GA and SA which later formed SAGA-FCM. PSO-LSSVM performed better than the other by considering R^2 and MRSE value.
- Wang et al. [54] also found that PSO-LSSVM outperformed other AI models. Again, they discovered that selected logs as an input estimate TOC batter than all logs as an input.
- In Johnson et al.'s [55] study ANN has been used for predicting geochemical well logs. ANN had produced high accuracy for TOC and S2 prediction where for S1 and HI the prediction accuracy was low.
- In Alizadeh et al. [56] work, ANN and $\Delta \log R$ were used for estimating TOC content from wireline logs. ANN provided higher precision compared to $\Delta \log R$ for estimating TOC content.
- An integrated hybrid neural network (IHNN) is developed in Zhu et al. [57] to work for TOC content estimation in Jiaoshiba area. From the result, it was observed that IHNN outperformed other AI models.

TABLE 8. Comparative studies of various AI methods done by previous researchers

<table>
<tr><th>Researchers</th><th>Study conducted</th><th>Study field</th><th>No. of data points</th><th>AI techniques compared</th><th>Performance criteria</th></tr>
<tr><td>[48]</td><td>TOC, S1 and S2 prediction</td><td>Bohay Bay Basin, China</td><td>125</td><td>CNN vs. BPANN</td><td><table>
<tr><td colspan="4">For selected logs as input (for TOC)</td></tr>
<tr><td colspan="2">R^2</td><td colspan="2">NRMSE</td></tr>
<tr><td>CNN</td><td>BP-ANN</td><td>CNN</td><td>BP-ANN</td></tr>
<tr><td>0.828</td><td>0.750</td><td>0.101</td><td>0.123</td></tr>
<tr><td colspan="4">For all logs as input (for TOC)</td></tr>
<tr><td colspan="2">R^2</td><td colspan="2">NRMSE</td></tr>
<tr><td>CNN</td><td>BP-ANN</td><td>CNN</td><td>BP-ANN</td></tr>
<tr><td>0.792</td><td>0.515</td><td>0.119</td><td>0.181</td></tr>
</table></td></tr>
<tr><td>[51]</td><td>TOC prediction</td><td>Beibu Gulf Basin, China</td><td>–</td><td>PSO-SVR vs. MLP</td><td><table>
<tr><td colspan="2">R^2</td></tr>
<tr><td>PSO-SVR</td><td>MLP-ANN</td></tr>
<tr><td>0.890</td><td>0.844</td></tr>
</table></td></tr>
<tr><td>[49]</td><td>TOC prediction</td><td>Barnett Shale, USA</td><td>800</td><td>TSK-FIS vs. M-FIS vs. FNN vs. SVM</td><td><table>
<tr><td colspan="4">R^2 (For training data)</td></tr>
<tr><td>TSK-FIS</td><td>M-FIS</td><td>FNN</td><td>SVM</td></tr>
<tr><td>0.937</td><td>0.926</td><td>0.876</td><td>0.871</td></tr>
<tr><td colspan="4">R^2 (For test data)</td></tr>
<tr><td>TSK-FIS</td><td>M-FIS</td><td>FNN</td><td>SVM</td></tr>
<tr><td>0.842</td><td>0.870</td><td>0.818</td><td>0.867</td></tr>
</table></td></tr>
<tr><td>[13]</td><td>TOC prediction</td><td>Barnett Shale, USA</td><td>689</td><td>SaDE-ANN vs. ANN</td><td><table>
<tr><td colspan="2">R^2</td></tr>
<tr><td>SaDE-ANN</td><td>ANN</td></tr>
<tr><td>0.98</td><td>0.88</td></tr>
</table></td></tr>
<tr><td>[54]</td><td>TOC prediction</td><td>Tonghua Basin, China</td><td>215</td><td>PSO-LSSVM vs. LSSVM vs. ANN-BP</td><td><table>
<tr><td colspan="3">R^2 (Train)</td></tr>
<tr><td>LSSVM</td><td>PSO-LSSVM</td><td>ANN-BP</td></tr>
<tr><td>0.9140</td><td>0.9273</td><td>0.9007</td></tr>
<tr><td colspan="3">R^2 (Test)</td></tr>
<tr><td>LSSVM</td><td>PSO-LSSVM</td><td>ANN-BP</td></tr>
<tr><td>0.9097</td><td>0.9205</td><td>0.8959</td></tr>
</table></td></tr>
<tr><td>[53]</td><td>TOC prediction</td><td>Ordos Basin, China</td><td>–</td><td>BP-ANN vs. LSSVM vs. PSO-LSSVM</td><td><table>
<tr><td colspan="3">R^2 (Train)</td></tr>
<tr><td>LSSVM</td><td>PSO-LSSVM</td><td>BP-ANN</td></tr>
<tr><td>0.9316</td><td>0.9451</td><td>0.9184</td></tr>
<tr><td colspan="3">RMSE</td></tr>
<tr><td>LSSVM</td><td>PSO-LSSVM</td><td>BP-ANN</td></tr>
<tr><td>0.4094</td><td>0.3383</td><td>0.5119</td></tr>
</table></td></tr>
<tr><td>[57]</td><td>TOC prediction</td><td>–</td><td>132</td><td>BP-Adaboost vs. KELM vs. SVM vs. IHNN</td><td><table>
<tr><td>Well A</td><td>BP-Adaboost</td><td>KELM</td><td>SVM</td><td>IHNN</td></tr>
<tr><td>RMAE</td><td>0.453</td><td>0.332</td><td>0.371</td><td>0.303</td></tr>
<tr><td>MSE</td><td>0.444</td><td>0.310</td><td>0.342</td><td>0.294</td></tr>
<tr><td>RRE</td><td>0.250</td><td>0.195</td><td>0.213</td><td>0.164</td></tr>
<tr><td>Well B</td><td>BP-Adaboost</td><td>KELM</td><td>SVM</td><td>IHNN</td></tr>
<tr><td>RMAE</td><td>0.542</td><td>0.547</td><td>0.695</td><td>0.453</td></tr>
<tr><td>MSE</td><td>0.586</td><td>0.670</td><td>0.865</td><td>0.442</td></tr>
<tr><td>RRE</td><td>0.355</td><td>0.523</td><td>0.485</td><td>0.284</td></tr>
</table></td></tr>
<tr><td>[59]</td><td>TOC prediction</td><td>Persian Gulf Basin, Iran</td><td>–</td><td>ANN vs. SVM</td><td><table>
<tr><td colspan="2">R^2</td></tr>
<tr><td>ANN</td><td>SVM</td></tr>
<tr><td>0.9077</td><td>0.9369</td></tr>
</table></td></tr>
<tr><td>[60]</td><td>TOC prediction</td><td>Sichuan Basin, China</td><td>185</td><td>ELM vs. ANN</td><td><table>
<tr><td colspan="4">R^2</td></tr>
<tr><td colspan="2">ELM</td><td colspan="2">ANN</td></tr>
<tr><td>Train</td><td>Test</td><td>Train</td><td>Test</td></tr>
<tr><td>0.868</td><td>0.854</td><td>0.937</td><td>0.931</td></tr>
</table></td></tr>
<tr><td>[64]</td><td>TOC prediction</td><td>South Pars Gas Field, Iran</td><td>2875</td><td>BP-ANN vs. TS-FIS</td><td><table>
<tr><td colspan="2">MSE</td></tr>
<tr><td>BP-ANN</td><td>TS-FIS</td></tr>
<tr><td>0.02</td><td>0.001</td></tr>
</table></td></tr>
</table>

- Bolandi et al. [59] have used fuzzy logic (FL), K-means clustering, ANN, and SVM in their work. They used FL and K-means clustering for searching the optimum pattern for estimating TOC. Then, they used SVM and ANN for estimating TOC from optimum well logs where SVM with RBF kernel outperformed ANN in terms of classification accuracy (0.9077 for ANN and 0.9369 for SVM) and reduced the computational time.
- Mahmoud et al. [14] have used ANN for developing an empirical equation for estimating TOC content from well logs. By using ANN weight and biases, this equation is developed. Then TOC was estimated with high accuracy for Barnett and Devonian shale formation where the developed equation was utilized.
- Extreme learning machine (ELM) has been used in Shi et al.'s [60] work. They used MLP-ANN for evaluating and comparing with ELM. They found that ELM can achieve high accuracy with maintaining high running speed.
- Tan et al. [61] have used support vector regression (SVR) technology for TOC content estimation from well logs. Different training algorithms in SVR which are Epsilon-SVR, Gaussian-SVR, RBF-SVR are used for determining the optimal algorithm in SVR. Then the optimal model of SVR is compared with $\Delta \log R$ model and empirical formulas where SVR outperformed the rest.
- In Ouadfeul and Aliouane's [62] study, the authors had used fuzzy logic and multilayer perceptron neural network with Levenberg Marquardt (MLP-ANN) for TOC content estimation from well logs. The result showed the power of MLP-ANN better than FL.
- MLP-ANN has been used in Ouadfeul and Aliouane's [63] work. MLP-ANN showed the efficiency for improving the shale gas reservoir characterization.
- Khoshnoodkia et al. [64] used $\Delta \log R$, ANN, FL, and Rock-eval for TOC content estimation. From the result, it is observed that these AI models were successful for predicting TOC content despite having poor source rock.

4.3. **Learning from the reviews.** The findings in relation to artificial intelligence applications in the TOC content estimation in shale gas reservoir characterization program are

- The numbers of research papers are indicative of the relevance and growth of artificial intelligence in shale gas reservoir characterization. Focusing on the contents of the available literature with information about various applications of AI in TOC content prediction, according to Figure 7 we can roughly estimate that about 52% papers are related to ANN, 15% are connected to support vector machines, about 13% − to hybrid intelligent systems, a little over 9% were dedicated to fuzzy logic, 7% to particle swarm optimization algorithm while about 4% was dedicated to the convolutional neural network and extreme learning machine.
- In this survey paper, researchers have used five countries as a study area for shale gas reservoir characterization which are respectively Australia, China, Egypt, Iran, and the USA. Among those five countries, China is used mostly as a study area by the researchers. Further, Barnett Shale from the USA is used most individually by the researchers as a study area.
- Artificial neural network (ANN) is the most common AI technique used by the researchers for predicting TOC content from well logs in shale gas reservoir where ANN is used for multiple inputs with respect to a single output. However, three authors in this survey carried their works by having multiple inputs and outputs.

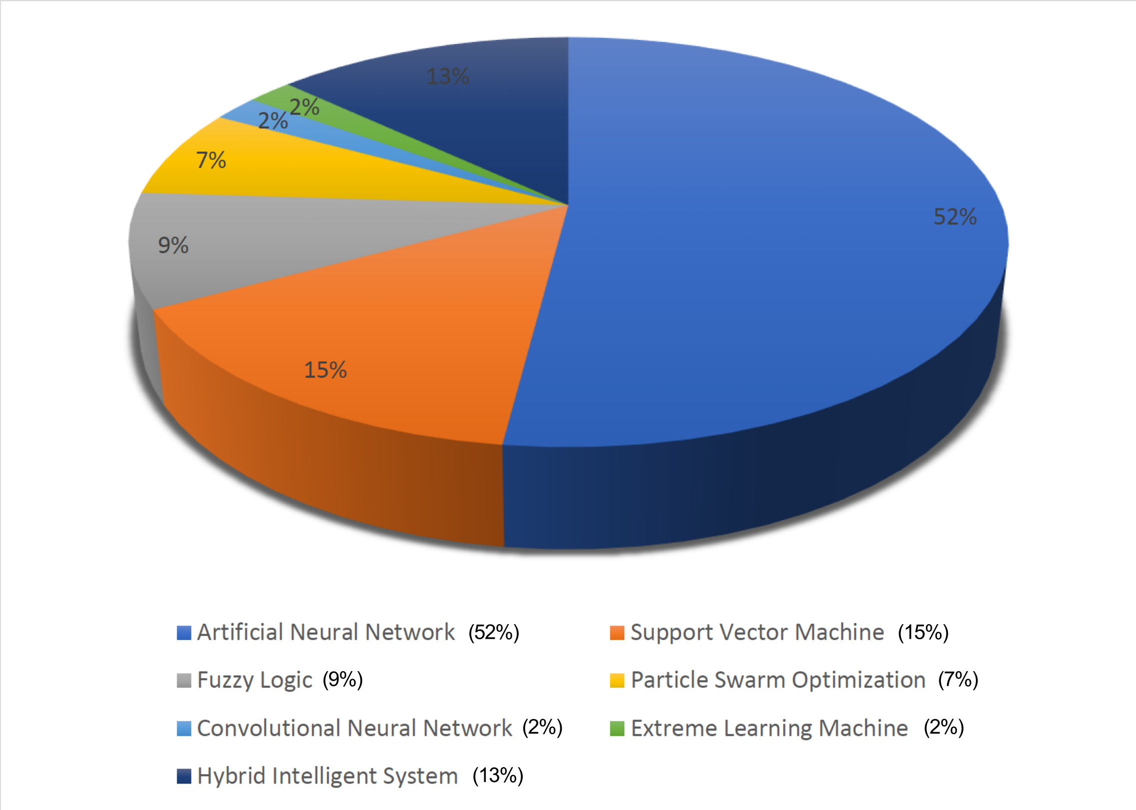

FIGURE 7. AI methods used in TOC estimation

- While reviewing the existing AI techniques, some AI techniques had come up with some limitations. These limitations had been handled and overcome in a robust way by the hybrid system which was proposed by some researchers.
- Hybrid models (containing SVM, FL, and ANN) optimized by DE or PSO, are found to be useful in estimating reservoir properties in shale gas reservoir.
- According to the "No Free lunch theorem", it is crucial to consider what AI techniques would be the best for predicting reservoir properties in different phenomena for shale gas reservoir characterization. Because AI techniques are unique with respect to adapt to new problems, obtain knowledge, deal with variability, etc.
- Important parameters need to be selected carefully among the plethora of data to decrease the size of datasets. This will increase the predictive performance of AI models. Additionally, the commanding importance of data preprocessing steps such as data normalization cannot be overemphasized given the boosting effect, they impart to the performance of AI models.

5. **Conclusion and Future Research.** This article reviews artificial intelligence technologies in shale gas field characterization and describes the latest research results and applications. Based on the above findings, the following conclusions can be drawn.

- This review briefly introduces the application of AI in characterizing oil shale oil and gas fields as an example of estimating TOC content from cable logs. The application of intelligent hybrid systems is also being investigated for development.
- Various AI techniques have been applied to estimate the characteristics of oil and gas shale fields from fixed line log curves such as TOC. Of these, artificial neural networks are the most used, and CNN and ELM are the least used.

- No single AI model can solve all problems. Therefore, a method can yield better results with one tool than with another. This article briefly addresses this issue in the "Strength and Limitation of Articial Intelligence Techniques" section.
- The study suggests that hybrid intelligence technology is the most successful and independent AI model, as it has the highest probability of estimating the properties of oil shale oil and gas fields (such as TOC) from cabled logs. A yet-to-be-explored optimization-based hybrid system can be investigated. These include artificial intelligent techniques such as ANN, SVM, and FL that can be optimized with other evolutionary algorithms such as GA, GP, and ES and other swarm intelligence such as GWO, ACO, and GSA to improve the learnability of these AI models.
- With the advancement of computing power, the future points in the direction of a more sophisticated deep learning system in TOC content prediction of shale gas reservoir characterization.
- AI can be used for estimating TOC from well logs of shale reservoirs which would provide a broad view for researchers. Therefore, there are many opportunities for future research.

Acknowledgments. The authors would like to thank and highly appreciate Petroleum Research Fund (PRF), Cost Center 0153AB-A33 and the project leader Dr. Eswaran Padmanabhan for supporting the research. The authors also would like to thank Universiti Teknologi PETRONAS (UTP) for Graduate Research Assistantship (GRA) scheme.

Nomenclature

AAPE	Average Absolute Percentage Error
ANFIS	Artificial Neuro-Fuzzy Intelligent System
AI	Artificial Intelligence
ANN	Artificial Neural Network
BPNN	Back Propagation Neural Network
BP-Adaboost	Back Propagation Adaboost
CNN	Convolutional Neural Network
DANN	Dynamic Artificial Neural Network
E&P	Exploration and Production
ELM	Extreme Learning Machine
FFBPN	Feed Forward Back Propagation Network
FIS	Fuzzy Intelligent Systems
FFNN	Feed Forward Neural Network
FFMLP	Feed Forward Multilayer Perceptron
GA	Genetic Algorithm
HIS	Hybrid Intelligent Systems
HI	Hydrogen Index
HML	Hybrid Machine Learning
IHNN	Integrated Hybrid Neural Network
KELM	Kernel Extreme Learning Machine
LSSVM	Least Square Support Vector Machine
MAE	Maximum Absolute Error
MLP	Multilayer Perceptron
MSE	Mean Square Error
MAPE	Mean Absolute Percentage Error
M-FIS	Mamdani Fuzzy Interference System
NFLTO	No Free Lunch Theorem for Optimization
NMSE	Normalized Mean Squared Error
OI	Oxygen Index

PSA	Particle Swarm Algorithm
PSO-ANN	Particle Swarm Optimization Artificial Neural Network
PSO-LSSVM	Particle Swarm Optimization Least Square Support Vector Machine
PSO	Particle Swarm Optimization
PSO-BP	Particle Swarm Optimization Back Propagation
R^2	Correlation Coefficient
RBNN	Radial Basis Neural Network
RBF	Radial Basis Function
RMSE	Root Mean Square Error
SaDE-ANN	Self-Adaptive Differential Evolution Artificial Neural Network
SAGA-FCM	Simulated Annealing Genetic Algorithm Fuzzy C-Means
SGS	Sequential Gaussian Simulation
SPE	Society of Petroleum Engineers
SVR	Support Vector Regression
SVM	Support Vector Machine
S2	The volume of hydrocarbons that formed during thermal pyrolysis of the sample
S1	The free hydrocarbons present in the sample before the analysis
S1+S2	Potential Yield
S3	The CO_2 yield during thermal breakdown of kerogen
TOC	Total Organic Carbon
TSK-FIS	Takagi-Sugeno-Kang Fuzzy Interference System

REFERENCES

[1] P. Wang, *Development and Applications of Production Optimization Techniques for Petroleum Fields*, Ph.D. Thesis, Stanford University Stanford, 2003.

[2] R. E. Gomory, Outline of an algorithm for integer solutions to linear programs, *Bull. Am. Math. Soc.*, vol.64, no.5, pp.275-278, 1958.

[3] A. H. Land and A. G. Doig, An automatic method for solving discrete programming problems, in *50 Years of Integer Programming 1958-2008*, Berlin, Springer, 2010.

[4] E. Mohagheghian, *An Application of Evolutionary Algorithms for WAG Optimisation in the Norne Field*, Ph.D. Thesis, Memorial University of Newfoundland, 2016.

[5] Z. Chen, *A Genetic Algorithm Optimizer with Applications to the SAGD Process*, Ph.D. Thesis, University of Calgary, 2013.

[6] H. Fujii and R. N. Horne, Multivariate optimization of networked production systems, *SPE Prod. Facil.*, vol.10, no.3, pp.165-171, 1995.

[7] A. T. Watson, J. H. Seinfeld, G. R. Gavalas and P. T. Woo, History matching in two-phase petroleum reservoirs, *Soc. Pet. Eng. J.*, vol.20, no.6, pp.521-532, 1980.

[8] E. Kamrani, Modeling and forecasting long-term natural gas (NG) consumption in Iran, using particle swarm optimization (PSO), *Computer Engineering*, 2010.

[9] H. Rahmanifard and T. Plaksina, Application of artificial intelligence techniques in the petroleum industry: A review, *Artif. Intell. Rev.*, vol.52, no.4, pp.2295-2318, 2019.

[10] J. McCarthy, From here to human-level AI, *Artif. Intell.*, vol.171, no.18, pp.1174-1182, 2007.

[11] S. Mohaghegh, Virtual-intelligence applications in petroleum engineering: Part 1 – Artificial neural networks, *J. Pet. Technol.*, vol.52, no.9, pp.64-73, 2000.

[12] G. Rable, Bloody engagements: John R. Kelso's Civil War, *Civ. War Book Rev.*, vol.19, no.3, p.11, 2017.

[13] S. Elkatatny, A self-adaptive artificial neural network technique to predict total organic carbon (TOC) based on well logs, *Arab. J. Sci. Eng.*, vol.44, no.6, pp.6127-6137, 2019.

[14] A. A. A. Mahmoud, S. Elkatatny, M. Mahmoud, M. Abouelresh, A. Abdulraheem and A. Ali, Determination of the total organic carbon (TOC) based on conventional well logs using artificial neural network, *Int. J. Coal Geol.*, vol.179, pp.72-80, 2017.

[15] S. Elkatatny and M. Mahmoud, Development of a new correlation for bubble point pressure in oil reservoirs using artificial intelligent technique, *Arab. J. Sci. Eng.*, vol.43, no.5, pp.2491-2500, 2018.

[16] I. Barman, A. Ouenes and M. Wang, Fractured reservoir characterization using streamline-based inverse modeling and artificial intelligence tools, *SPE Annual Technical Conference and Exhibition*, Dallas, Texas, 2000.

[17] S. Mohaghegh, R. Arefi, S. Ameri and M. H. Hefner, A methodological approach for reservoir heterogeneity characterization using artificial neural networks, *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, 1994.

[18] A. Al-AbdulJabbar, S. Elkatatny, M. Mahmoud, K. Abdelgawad and A. Al-Majed, A robust rate of penetration model for carbonate formation, *J. Energy Resour. Technol.*, vol.141, no.4, 2019.

[19] K. Abdelgawad, S. Elkatatny, T. Moussa, M. Mahmoud and S. Patil, Real-time determination of rheological properties of spud drilling fluids using a hybrid artificial intelligence technique, *J. Energy Resour. Technol.*, vol.141, no.3, 2019.

[20] S. Elkatatny, Real-time prediction of rheological parameters of KCL water-based drilling fluid using artificial neural networks, *Arab. J. Sci. Eng.*, vol.42, no.4, pp.1655-1665, 2017.

[21] S. Elkatatny, Application of artificial intelligence techniques to estimate the static Poisson's ratio based on wireline log data, *J. Energy Resour. Technol.*, vol.140, no.7, 2018.

[22] A. A. Mahmoud, S. Elkatatny, A. Ali and T. Moussa, Estimation of static young's modulus for sandstone formation using artificial neural networks, *Energies*, vol.12, no.11, 2019.

[23] A. A. Mahmoud, S. Elkatatny, W. Chen and A. Abdulraheem, Estimation of oil recovery factor for water drive sandy reservoirs through applications of artificial intelligence, *Energies*, vol.12, no.19, 2019.

[24] Y. Wang and S. Salehi, Application of real-time field data to optimize drilling hydraulics using neural network approach, *J. Energy Resour. Technol.*, vol.137, no.6, 2015.

[25] A. S. Ahmed, A. A. Mahmoud and S. Elkatatny, Fracture pressure prediction using radial basis function, *Proc. of the 2019 AADE National Technical Conference and Exhibition*, Denver, CO, USA, pp.9-10, 2019.

[26] S. Ahmed, A. A. Mahmoud, S. Elkatatny, M. Mahmoud and A. Abdulraheem, Prediction of pore and fracture pressures using support vector machine, *International Petroleum Technology Conference*, 2019.

[27] D. A. Al-Shehri, Oil and gas wells: Enhanced wellbore casing integrity management through corrosion rate prediction using an augmented intelligent approach, *Sustainability*, vol.11, no.3, p.818, 2019.

[28] S. Salehi, G. Hareland, K. K. Dehkordi, M. Ganji and M. Abdollahi, Casing collapse risk assessment and depth prediction with a neural network system approach, *J. Pet. Sci. Eng.*, vol.69, nos.1-2, pp.156-162, 2009.

[29] M. Ahmad, O. Iqbal and A. A. Kadir, Quantification of organic richness through wireline logs: A case study of Roseneath shale formation, Cooper basin, Australia, *IOP Conference Series: Earth and Environmental Science*, vol.88, 2017.

[30] Q. R. Passey, K. Bohacs, W. L. Esch, R. Klimentidis and S. Sinha, From oil-prone source rock to gas-producing shale reservoir-geologic and petrophysical characterization of unconventional shale gas reservoirs, *CPS/SPE International Oil & Gas Conference and Exhibition in China*, Beijing, China, 2010.

[31] C. H. Sondergeld, K. E. Newsham, J. T. Comisky, M. C. Rice and C. S. Rai, Petrophysical considerations in evaluating and producing shale gas resources, *SPE Unconventional Gas Conference*, Pittsburgh, Pennsylvania, USA, 2010.

[32] Y. Altowairqi, R. Rezaee, B. Evans and M. Urosevic, Shale elastic property relationships as a function of total organic carbon content using synthetic samples, *J. Pet. Sci. Eng.*, vol.133, pp.392-400, 2015.

[33] S. L. Montgomery, D. M. Jarvie, K. A. Bowker and R. M. Pollastro, Mississippian Barnett shale, fort worth basin, north-central Texas: Gas-shale play with multi-trillion cubic foot potential, *AAPG Bull.*, vol.89, no.2, pp.155-175, 2005.

[34] D. J. Ross and R. M. Bustin, Impact of mass balance calculations on adsorption capacities in microporous shale gas reservoirs, *Fuel*, vol.86, nos.17-18, pp.2696-2706, 2007.

[35] J. Ding, X. Chen, X. Jiang, W. Bin and J. Zhang, Application of AVF inversion on shale gas reservoir TOC prediction, *Proc. of the SEG Annual Meeting: Society of Exploration Geophysicists*, New Orleans, LA, USA, 2015.

[36] P. Wang, Z. Chen, X. Pang, K. Hu, M. Sun and X. Chen, Revised models for determining TOC in shale play: Example from Devonian Duvernay Shale, Western Canada Sedimentary Basin, *Mar. Pet. Geol.*, vol.70, pp.304-319, 2016.

[37] J. W. Schmoker, Determination of organic content of appalachian devonian shales from formation-density logs: Geologic notes, *AAPG Bull.*, vol.63, no.9, pp.1504-1509, 1979.

[38] J. W. Schmoker and T. C. Hester, Organic carbon in Bakken Formation, United States portion of Williston basin, *AAPG Bull.*, vol.67, no.12, pp.2165-2174, 1983.

[39] Q. R. Passey, S. Creaney, J. B. Kulla, F. J. Moretti and J. D. Stroud, A practical model for organic richness from porosity and resistivity logs, *AAPG Bull.*, vol.74, no.12, pp.1777-1794, 1990.

[40] F. M. Chang, Characteristics analysis for small data set learning and the comparison of classification methods, *Proc. of the 7th WSEAS International Conference on ARTIFICIAL INTELIGENCE, KNOWLEDGE ENGINEERING and Data Bases (AIKED'08)*, pp.23-25, 2008.

[41] T. Helmy, A. Fatai and K. Faisal, Hybrid computational models for the characterization of oil and gas reservoirs, *Expert Syst. Appl.*, vol.37, no.7, pp.5353-5363, 2010.

[42] F. A. Anifowose and A. Abdulraheem, Prediction of porosity and permeability of oil and gas reservoirs using hybrid computational intelligence models, *Society of Petroleum Engineers North Africa Technical Conference and Exhibition*, 2010.

[43] J. M. Mendel, Type-2 fuzzy sets and systems: An overview, *IEEE Comput. Intell. Mag.*, vol.2, no.1, pp.20-29, 2007.

[44] V. Van Belle, K. Pelckmans, S. Van Huffel and J. A. Suykens, Improved performance on high-dimensional survival data by application of survival-SVM, *Bioinformatics*, vol.27, no.1, pp.87-94, 2011.

[45] D. H. Wolpert and W. G. Macready, No free lunch theorems for optimization, *IEEE Trans. Evol. Comput.*, vol.1, no.1, pp.67-82, 1997.

[46] Y. Chen and Y. Zhao, A novel ensemble of classifiers for microarray data classification, *Appl. Soft Comput.*, vol.8, no.4, pp.1664-1669, 2008.

[47] T. Helmy, M. M. Al-Harthi and M. T. Faheem, Adaptive ensemble and hybrid models for classification of bioinformatics datasets, *Trans. Fuzzy Neural Netw Bioinform Glob J. Technol. Optim.*, vol.3, pp.20-29, 2012.

[48] H. Wang, W. Wu, T. Chen, X. Dong and G. Wang, An improved neural network for TOC, S1 and S2 estimation based on conventional well logs, *J. Pet. Sci. Eng.*, 2019.

[49] A. A. Mahmoud, S. Elkatatny, A. Z. Ali, M. Abouelresh and A. Abdulraheem, Evaluation of the total organic carbon (TOC) using different artificial intelligence techniques, *Sustainability*, vol.11, no.20, 2019.

[50] M. R. Shalaby, N. Jumat, D. Lai and O. Malik, 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.*, vol.176, pp.369-380, 2019.

[51] J. Rui, H. Zhang, D. Zhang, F. Han and Q. Guo, Total organic carbon content prediction based on support-vector-regression machine with particle swarm optimization, *J. Pet. Sci. Eng.*, vol.180, pp.699-706, 2019.

[52] Y. A. Nezhad, A. Moradzadeh and M. R. Kamali, A new approach to evaluate organic geochemistry parameters by geostatistical methods: A case study from western Australia, *J. Pet. Sci. Eng.*, vol.169, pp.813-824, 2018.

[53] P. Wang and S. Peng, A new scheme to improve the performance of artificial intelligence techniques for estimating total organic carbon from well logs, *Energies*, vol.11, no.4, DOI: 10.3390/en11040747, 2018.

[54] P. Wang, S. Peng and T. He, A novel approach to total organic carbon content prediction in shale gas reservoirs with well logs data, Tonghua Basin, China, *J. Nat. Gas Sci. Eng.*, vol.55, pp.1-15, 2018.

[55] L. M. Johnson, R. Rezaee, A. Kadkhodaie, G. Smith and H. Yu, Geochemical property modelling of a potential shale reservoir in the Canning Basin (Western Australia), using artificial neural networks and geostatistical tools, *Comput. Geosci.*, vol.120, pp.73-81, 2018.

[56] B. Alizadeh, K. Maroufi and M. H. Heidarifard, Estimating source rock parameters using wireline data: An example from Dezful Embayment, South West of Iran, *J. Pet. Sci. Eng.*, vol.167, pp.857-868, 2018.

[57] L. Zhu et al., Prediction of total organic carbon content in shale reservoir based on a new integrated hybrid neural network and conventional well logging curves, *J. Geophys. Eng.*, vol.15, no.3, pp.1050-1061, 2018.

[58] P. Tahmasebi, F. Javadpour and M. Sahimi, Data mining and machine learning for identifying sweet spots in shale reservoirs, *Expert Syst. Appl.*, vol.88, pp.435-447, 2017.

[59] V. Bolandi, A. Kadkhodaie and R. Farzi, Analyzing organic richness of source rocks from well log data by using SVM and ANN classifiers: A case study from the Kazhdumi Formation, the Persian Gulf basin, offshore Iran, *J. Pet. Sci. Eng.*, vol.151, pp.224-234, 2017.
[60] X. Shi, J. Wang, G. Liu, L. Yang, X. Ge and S. Jiang, Application of extreme learning machine and neural networks in total organic carbon content prediction in organic shale with wire line logs, *J. Nat. Gas Sci. Eng.*, vol.33, pp.687-702, 2016.
[61] M. Tan, X. Song, X. Yang and Q. Wu, Support-vector-regression machine technology for total organic carbon content prediction from wireline logs in organic shale: A comparative study, *J. Nat. Gas Sci. Eng.*, vol.26, pp.792-802, 2015.
[62] S.-A. Ouadfeul and L. Aliouane, Total organic carbon prediction in shale gas reservoirs using the artificial intelligence with a comparative study between fuzzy logic and neural network, *The 14th International Congress of the Brazilian Geophysical Society & EXPOGEF*, Rio de Janeiro, Brazil, pp.1390-1393, 2015.
[63] S.-A. Ouadfeul and L. Aliouane, Shale gas reservoirs characterization using neural network, *Energy Procedia*, vol.59, pp.16-21, 2014.
[64] M. Khoshnoodkia, H. Mohseni, O. Rahmani and A. Mohammadi, TOC determination of Gadvan Formation in South Pars Gas field, using artificial intelligent systems and geochemical data, *J. Pet. Sci. Eng.*, vol.78, no.1, pp.119-130, 2011.
[65] P. J. Braspenning, F. Thuijsman and A. J. M. M. Weijters, *Artificial Neural Networks: An Introduction to ANN Theory and Practice*, Springer Science & Business Media, 1995.
[66] L. A. Zadeh, Fuzzy sets, *Inf. Control*, vol.8, no.3, pp.338-353, 1965.
[67] T. Terano, K. Asai and M. Sugeno, *Applied Fuzzy Systems*, Academic Press, 2014.
[68] T. J. Ross, J. M. Booker and A. C. Montoya, New developments in uncertainty assessment and uncertainty management, *Expert Syst. Appl.*, vol.40, no.3, pp.964-974, 2013.
[69] R. Eberhart and J. Kennedy, A new optimizer using particle swarm theory, *Proc. of the 6th International Symposium on Micro Machine and Human Science*, pp.39-43, 1995.
[70] V. Vapnik, Pattern recognition using generalized portrait method, *Autom. Remote Control*, vol.24, pp.774-780, 1963.
[71] C. Cortes and V. Vapnik, Support-vector networks, *Mach. Learn.*, vol.20, no.3, pp.273-297, 1995.
[72] J. Nayak, B. Naik and H. Behera, A comprehensive survey on support vector machine in data mining tasks: Applications & challenges, *Int. J. Database Theory Appl.*, vol.8, no.1, pp.169-186, 2015.
[73] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer and D. Haussler, Support vector machine classification and validation of cancer tissue samples using microarray expression data, *Bioinformatics*, vol.16, no.10, pp.906-914, 2000.
[74] J. A. Suykens and J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.*, vol.9, no.3, pp.293-300, 1999.
[75] L. Guan, J. Wang and L. Wang, Hybrid computational intelligence based kinematic analysis for parallel machine tool, *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme – System Security and Assurance (Cat. No. 03CH37483)*, Washington, DC, pp.2763-2768, 2003.
[76] A. Tsakonas and G. Dounias, Hybrid computational intelligence schemes in complex domains: An extended review, *Proc. of the Second Hellenic Conference on AI: Methods and Applications of Artificial Intelligence*, pp.494-512, 2002.
[77] A. Chandra and X. Yao, Evolving hybrid ensembles of learning machines for better generalisation, *Neurocomputing*, vol.69, nos.7-9, pp.686-700, 2006.
[78] M. Khashei, R. F. Mokhatab, M. Bijari and S. R. Hejazi, A hybrid computational intelligence model for foreign exchange rate forecasting, 2011.
[79] E.-N. Dragoi, S. Curteanu, A.-I. Galaction and D. Cascaval, Optimization methodology based on neural networks and self-adaptive differential evolution algorithm applied to an aerobic fermentation process, *Appl. Soft Comput.*, vol.13, no.1, pp.222-238, 2013.
[80] F. A. Anifowose, J. Labadin and A. Abdulraheem, Hybrid intelligent systems in petroleum reservoir characterization and modeling: The journey so far and the challenges ahead, *J. Pet. Explor. Prod. Technol.*, vol.7, no.1, pp.251-263, 2017.