Energy Reports 8 (2022) 6551–6562

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

Data driven models to predict pore pressure using drilling and petrophysical data

Farshad Jafarizadeh [a], Meysam Rajabi [b], Somayeh Tabasi [c], Reza Seyedkamali [d], Shadfar Davoodi [e,*], Hamzeh Ghorbani [f,g,**], Mehdi Ahmadi Alvar [h], Ahmed E. Radwan [i,***], Mako Csaba [j]

[a] Department of Petroleum Engineering, Amirkabir University of Technology, Tehran, Iran
[b] Department of Mining Engineering, Birjand University of Technology, Birjand, Iran
[c] Faculty of Industry and Mining, University of Sistan and Baluchestan, Khash, Zahedan, Iran
[d] Department of Petroleum Engineering, Amir Kabir University, Tehran, Iran
[e] School of Earth Sciences & Engineering, Tomsk Polytechnic University, Lenin Avenue, Tomsk, Russia
[f] Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
[g] Faculty of General Medicine, University of Traditional Medicine of Armenia (UTMA) - 38a Marshal Babajanyan St., Yerevan 0040, Armenia
[h] Faculty of Engineering, Department of Computer Engineering, Shahid Chamran University, Ahwaz, Iran
[i] Faculty of Geography and Geology, Institute of Geological Sciences, Jagiellonian University, Gronostajowa 3a, 30-387, Kraków, Poland
[j] Institute of the Information Society, University of Public Service, 1083 Budapest, Hungary

ARTICLE INFO

Article history:
Received 1 January 2022
Received in revised form 22 April 2022
Accepted 27 April 2022
Available online xxxx

Keywords:
K-nearest neighbor distance weighted (DWKNN)
Pore pressure prediction
Hybrid machine learning
Feature selection
Root mean squared error

ABSTRACT

The mud weight window (MW) determination is one of the most important parameters in drilling oil and gas wells, where accurate design can secure the drilled well and deliver a stable borehole. In this paper, novel algorithms based on the most influential set of input features are developed to predict pore pressure, including rate of penetration (ROP), deep resistivity (ILD), density (RHOB), photoelectric index (PEF), corrected gamma ray (CGR), compression-wave velocity (Vp), weight on bit (WOB), shear-wave velocity (Vs) and pore compressibility (Cp). The algorithms used in this study are as follows: 1) machine learning algorithms (ML), these are the K-nearest neighbor (KNN) algorithm, weighted K-Nearest Neighbor (WKKNN), and distance weighted KNN (DWKNN); 2) hybrid machine learning algorithms (HML), which include the combination of three ML with particle swarm optimization (PSO) (KNN-PSO, WKNN-PSO and DWKNN-PSO). The 2875-record dataset used in this study was collected from three wells (S1, S2 and S3) in one of the gas reservoirs (Tabnak field) in Iran. After comparing the performance accuracy of all algorithms, DWKNN-PSO has the best performance accuracy compared to other algorithms presented in this paper (for the total dataset of wells S1 and S2: R^2 = 0.9656 and RMSE = 12.6773 psi). Finally, the generalizability of the best predictive algorithm for PP, DWKNN-PSO, is evaluated by testing the proposed algorithm on an unseen dataset from another well (S3) in the field under study, where the DWKNN-PSO algorithm provides PP predictions in well S3 with high accuracy, R2 = 0.9765 and RMSE = 9.7545 psi, confirming its ability to be used in PP prediction in the studied field.

1. Introduction

One of the requirements for achieving hydrocarbon production from deep underground is the use of exploration operations and drilling for oil and gas wells (Elkatatny et al., 2020; Rashidi et al., 2020). Therefore, drilling operations are very important and significant in the oil and gas industry (Hazbeh et al., 2021a). During drilling operations in sedimentary succession, different pressures are encountered in these sedimentary formations. These different pressures must be identified in such a way that allows for determining the drilling mud pressure and determining the most appropriate drilling mud window. The accurate determination of

* Corresponding author at: School of Earth Sciences & Engineering, Tomsk Polytechnic University, Lenin Avenue, Tomsk, Russia.
** Corresponding author at: Faculty of General Medicine, University of Traditional Medicine of Armenia (UTMA) - 38a Marshal Babajanyan St., Yerevan 0040, Armenia.
*** Corresponding author at: Faculty of Geography and Geology, Institute of Geological Sciences, Jagiellonian University, Gronostajowa 3a, 30-387, Kraków, Poland.
E-mail addresses: davoodis@hw.tpu.ru (S. Davoodi), hamzehghorbani68@yahoo.com (H. Ghorbani), radwanae@yahoo.com (A.E. Radwan).

https://doi.org/10.1016/j.egyr.2022.04.073

the drilling mud window can prevent problems such as loss and blowout, which have a direct impact on drilling costs (Abdali et al., 2021; Abad et al., 2021a,b; Jia et al., 2021; Radwan et al., 2019, 2020; Radwan, 2021; Abdelghany et al., 2021). Pore pressure (PP) is the fluid pressure in the pore space, and when it exceeds the hydrostatic pressure, an overpressure situation occurs (Walsh, 1981; Radwan et al., 2019, 2020; Radwan, 2021). One of the key solutions for safe well drilling is the identification and understanding of PP and fracture pressure (FP) in the subsurface, which includes wellbore planning, analyzes of wellbore stability, casing design, drilling fluid designs, drilling plans, structure optimization, and production optimization (Brainard, 2006; Hu et al., 2013; Yin et al., 2020; Zhang et al., 2022). In addition to selectively selecting injection and production and determining a hydrocarbon migration route, accurate determination of PP aids in estimating the location of pressure problems (Hou et al., 2017; Keshavarzi and Jahanbakhshi, 2013; Naveshki et al., 2021).

In general, PP is divided into normal pressure (pressure gradient of 0.433 psi/ft) and abnormal pressure (overpressure higher than normal pressure due to an additional pressure source and under-pressure due to production) (Radwan et al., 2019, 2020; Radwan, 2021; Radwan and Sen, 2021a,b). Abnormal pressure causes problems such as adhesion of various pipes and severe mud loss (Bahrami et al., 2020; Du et al., 2020; Kremieniewski, 2020). Accurate real-time PP determination and prediction can be demonstrated by well drilling path prediction, drilling mud schedule, and well stability analysis, and can reduce drilling time and cost (Baouche et al., 2020; Dawson et al., 2020). Therefore, many researchers have tried to provide empirical equations to predict this important parameter, which include:

In 1943, Terzaghi et al. defined overburden pressure based on PP and effective pressure (Terzaghi, 1943). As shown in Eq. (1):

$$P_{over} = PP + P_{eff} \tag{1}$$

where;

P_{over} is overburden pressure; PP is pore pressure and P_{eff} is effective pressure.

Later, in 1975, Biot and Willis proposed an experimental relationship between overburden pressure, effective stress, and PP with a coefficient called the Biot coefficient (Eq. (2)). The Biot coefficient is the relative coefficient of change in the pore fluid volume relative to the volume of the whole rock, and if this liquid is free and does not prevent it from leaving the rock, its value is equal to one. It should be noted that this parameter is valid only for homogeneous rocks, and this theory (Biot coefficient) is not true for heterogeneous rocks (Biot and Willis, 1957).

$$PP = \frac{P_{over} - P_{eff}}{\omega} \tag{2}$$

where;

PP is pore pressure; P_{over} is overburden pressure; P_{eff} is effective pressure and ω is Biot coefficient.

Eaton proposed two formulas for calculating the PP using the two parameters (sonic log and gamma ray log) in 1975, which are presented in Eqs. (3) and (4).

$$PP = P_{over} - (P_{over} - P_{hyd})\left(\frac{\Delta_{ts}}{\Delta_t}\right)^x \tag{3}$$

$$PP = P_{over} - (P_{over} - P_{hyd})\left(\frac{GR_s}{GR}\right)^x \tag{4}$$

where;

PP is pore pressure; P_{over} is overburden pressure; P_{hyd} is hydrostatic pressure; Δ_{t} is a sonic log; Δ_{ts} is a sonic log in shale; GR is a gamma ray log; GR_{s} is a gamma ray log in shale and x is Eaton coefficient.

Much research has been conducted in the last decade to develop more accurate PP predictions using machine learning. In 2013, Nour and AlBinHassan predicted the PP in an oil field in northwest Saudi Arabia using the support vector machine (SVM) algorithm regression method. They used seismic data to predict this important parameter. Finally, they concluded that this algorithm is capable of predicting PP with high performance accuracy (Nour and AlBinHassan, 2013).

In 2014, Abidin used an artificial neural network (ANN) algorithm to predict the PP in fields characterized by normal and abnormal pressure. This method consists of three layers, where the optimal number of inputs for the input, hidden, and output layers is 2, 10, and 1, respectively. The function used in this algorithm is of the sigmoid function type. The performance accuracy of this algorithm is very high and is about 5.0048% (Abidin, 2014).

One year later, Aliouane et al. (2015) predicted the PP parameter in horizontal wells drilled in shale gas formations using fuzzy logic (FL) and the multilayer perceptron neural network (MLP-ANN). In this paper (Aliouane et al., 2015), the input data used to predict PP are natural gamma-ray (GR), neutron porosity (NPHI), slowness of compression (SC), and shear wave velocity (VS).

Three years later, Kiss et al. (2018) predicted the PP parameter using an ANN algorithm based on two important parameters: mechanical specific energy drilling (MSE) and drilling efficiency (DE). Their field of study was these wellbores that were drilled in an Iranian sandstone formation. The results showed that ANN provides more reliable results for PP prediction (Kiss et al., 2018).

Ahmed et al. (2019) predicted the PP parameter using the ANN, radial basis function (RBF), FL, support vector machine (SVM), and functional networks (FN) algorithms. They have used formation density (RHOB), porosity (ϕ), compression time (Δt), predicted weight on bit (WOB), rotary speed (RPM), penetration rate (ROP) and mud density (MW). The results of this study showed that the SVM has a shorter execution time and higher performance accuracy than other methods. The performance accuracy of this method was $R^2 = 0.9950$ and ARE = 0.14%.

Andrian et al. (2020) predicted the PP parameter using adaptive neuro fuzzy inference system (ANFIS) algorithms, which is a combination of ANN and FIS machine learning and is based on 2D seismic data parameters. The results show that the ANFIS algorithm can show PP with 70% performance accuracy.

The PP parameter was predicted by (Farsi et al., 2021b) using a combination of least squares support-vector machine (LSSVM), multilayer perceptron (MLP) and extreme learning machine (ELM) with particle swarm optimization (PSO). In this article, a 1972 dataset from Iranian oil field has been used. The results show that MELM–PSO has highest performance accuracy (RMSE = 11.551 psi).

Radwan et al. (2022) used 25,935 data records to predict the PP parameter in the Mangahewa gas field, New Zealand. They applied several machine learning ML techniques, including the Extreme learning machine (ELM), multi-layer perceptron (MLP), multi-linear regression (MLR, with gradient descent optimizer), optimizer formula (OF, polynomial equation fit with optimizer), random forest (RF), support vector regression (SVR), Adaboost (ADA, a boosted decision tree (DT) model), DT, and transparent open box (TOB). The results show that the most accurate models for predicting PP are decision tree (DT), adaboost (ADA), random forest (RF) and transparent open box (TOB).

In this paper, using a 2875 dataset that represents three wells from Iranian gas fields located in southern Iran, we have applied a new model for PP prediction using three new techniques. These are the K-nearest neighbor (KNN), weight K-nearest neighbor (WKNN), and distance weight K-nearest neighbor (DWKNN), and these models are combined with a PSO algorithm, which has not been used in any of the previous PP prediction published articles. The purpose of this article is to apply the three new artificial

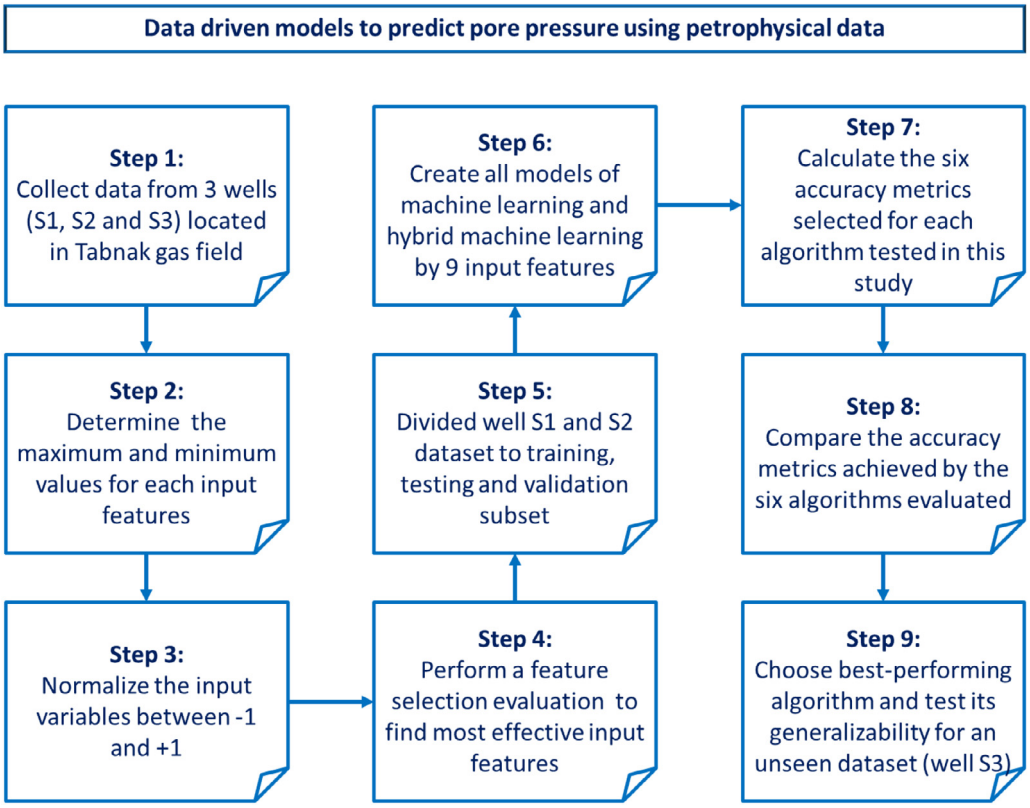

Fig. 1. Illustration of workflow chart used to develop and test six ML and HML algorithms for prediction of PP.

intelligence techniques to the studied gas field data records and compare them with each other. Moreover, in this study, we used two novel parameters (WOB and ROP) and the result shows that these algorithms have an effect on predicted PP. The WOB is inversely related to PP in exchange for increasing depth, where it decreases for every PP increase. And according to the definition of ROP, this parameter is directly related to WOB, so both WOB and ROP are related (inversely) to PP. This paper aims to achieve an important (costly to access) parameter using routine drilling parameters. Therefore, this article introduces two important drilling parameters and their tire emphasizes the formation pressure, which is indeed a new contribution in the method proposed for predicting PP. Because they are simple and fast to implement, the KNN-based algorithms can be used to evaluate datasets with noisy or incomplete data samples. The main concerns with the KNN based algorithms are the negative effects of K values and neighbor weights on the accuracy of the predicted target value. As a result, the PSO optimizer was employed in this study to estimate optimal weights and K values to enhance the model's prediction accuracy.

2. Methodology

2.1. Workflow chart

The diagram shown in Fig. 1 shows the sequence of construction and evaluation steps of ML and HML algorithms in order to predict PP and the best performance accuracy for the algorithms. Initially, data was collected from three wells, namely: S1, S2 and S3 from the Iranian gas field. Then, we set the minimum and maximum values for each feature in order to normalize them. To normalize the data in this research, we used the normalization equation in Eq. (5) to be in the range of -1 and $+1$. Further, we used the feature selection to select the effective input variables (9 input features). In the next step, we classify the data into two parts: training and testing. Then we compare the ML and HML algorithms using statistical metrics. Finally we generalize the best algorithm using well S3, which is related to the same field to determine that this algorithm can be used in other wells.

$$S_i^l = \left(\frac{S_i^l - Smin^l}{Smax^l - Smin^l}\right) * 2 - 1 \tag{5}$$

where:

S_i^l = the value of data record I;

$Smin^l$ = the minimum value of the data records; and, $Smax^l$ = the maximum value of all the data records.

2.2. Selection of feature for predicting pore pressure

Feature selection is a technique used for the determination of the optimal number of input variables to include in the machine learning models. In this study, an MLP-GA algorithm was deployed to identify the most influential set of input variables by applying a multiple tarin and test sublets of data points. An analysis of trial-and-error was performed to discover the most effective model for feature selection. The MLP model of two layers containing five and six neurons in its 1st and 2nd hidden layers was identified as the most effective structure for feature selection at RMSE minimization of the PP value predicted.

First, the input variables were divided into different combinations of variables, and the minimum value of RMSE was considered as the evolution criteria in selecting the influential features. Then, the PP was computed by the feature selection model. To model the PP, the MLP-GA and the combination of input variables with the most effective parameters that were found to perform the best predictions in terms of accuracy were chosen.

2.3. Predictive algorithms

Machine learning modeling has received a great deal of attention from researchers in various areas of science and engineering over the last decade (Abad et al., 2021b; Ghorbani et al., 2017, 2020, 2019; Kombo et al., 2020; Ranaee et al., 2021). In this regard, a variety of machine learning algorithms have been widely used in earth sciences and petroleum engineering to tackle different classification and regression problems (Abad et al., 2022, 2021a,b; Barjouei et al., 2021; Hazbeh et al., 2021a,b; Rajabi et al., 2021; Torabi et al., 2019). In this study, the K-Nearest Neighbor (KNN) algorithm and two of its improved versions, weighted K-Nearest Neighbor (WKKNN) and distance weighted KNN (DW-KNN) are applied for predicting pore pressure via conventional well logging data.

2.3.1. KNN algorithm

The KNN algorithm is a simple yet robust machine learning (ML) technique that is used for tackling classification and regression problems. In the KNN technique, data points that are not trained can be approximated using their proximity to data points that are trained (Navot et al., 2005; Rashidi et al., 2020). One of the KNN features that distinguishes it from other ML methods is its capability to deal with incomplete observations and noisy data (Brajard et al., 2020; Nguyen et al., 2019). This approach can identify the most influential areas from the noisy data records. Since field data recorded by well logging tools is commonly noisy, so the KNN based algorithms can be considered an appropriate solution to accurately medal pore pressure. Euclidean, Minkowski, Chebyshev, and Manhattan metrics can be used for machining of points performed in respect of the distance measured. Consider X and Y to be two sets with n points each, where $X = (x_1, x_2, x_3, \ldots, x_n)$, $Y = (y_1, y_2, y_3, \ldots, y_n)$, and $(i = 1, 2, 3, 4, I, n)$. Then, the distance between nearby points and the intended point can be enumerated. The following relationship (Eq. (6)) defines the distance between the closest points and the point desired (Kombo et al., 2020).

$$D_i = \sum_{i=1}^{n} [(|x_i - y_i|^2)]^{1/2} \tag{6}$$

where;

D_i stand for the calculated Euclidean distance between data sample X and the data sample i of the training set, x_i represents the value of test sample, y_i denote the value of ith training sample. In order to predict the target value using KNN, the following steps are followed:

1. Determine the distance between a new sample and the neighboring points by applying Equation one.
2. Sort all the distance values determined in the first step in increasing order.
3. Find the optimal K value based on an accuracy metric using an optimization technique.
4. Compute an inverse distance mean by the K neighboring samples using Eq. (7).

$$C_P = \frac{1}{K} \sum_{t=1}^{K} C_t \tag{7}$$

where;

C_P denotes the predicted value of output variable for test data sample, C_t stands for the values of output variable of the nearest neighbor t.

2.3.2. WKNN and DW-KNN algorithms

The WKKN technique is an improved version of the KNN algorithm, where a weight is assigned to each matching K data record based on its distance to the other k-1 nearest adjacent records from the test data. In this method, a lower weight is assigned to the neighbors that have a longer distance to the test data record than those closest neighbors. As a result, the nearest neighbors have the highest degree of influence on the predicted target value. The weight function for WKNN can be formally defined as shown below (Eq. (8)) (Farsi et al., 2021a; Huang et al., 2017).

$$\omega_i = \begin{cases} \frac{d_k - d_1}{d_k - d_1}, & d_k \neq d_1 \\ 1, & d_k = d_1 \end{cases} \tag{8}$$

Applying such weights to determine of each neighbor's record influence helps the WKNN in overcoming the negative effect of neighborhood K. The DWKNN is an extended version of the WKNN that was proposed to further overcome the neighborhood K negative influence. In the DW-KNN, the linear mapping of the WKNN is extended. The following relationship defines the dual-weighted function for the DW-KNN algorithm (see Eq. (9)) (Huang et al., 2017).

$$\omega_i = \begin{cases} \frac{d_k - d_1}{d_k - d_1} \cdot \frac{d_k + d_1}{d_k + d_1}, & d_k \neq d_1 \\ 1, & d_k = d_1 \end{cases} \tag{9}$$

For the DW-KNN and WKNN, the dependent variable value (C_{un}) can be predicted using Eq. (10), in which W_i represents the weight of dependent variable to be utilized for neatest neighbor i.

$$C_{un} = \sum_{i=1}^{k} w_i C_i \tag{10}$$

The particle swarm optimization algorithm is used in this study to estimate the optimal value for K and to optimize the set of weights assigned to each variable in the dataset under consideration.

2.3.3. Optimization algorithm

Particle swarm optimization (PSO) is an evolutionary computation method that was developed on the basis of various patterns of populations existing in nature, such as fish, birds, and insects (Eberhart and Kennedy, 1995; Rashidi et al., 2021). This technique solves the optimization problems by promoting initial populations where the solutions are called "particles". In this approach, a swarm is represented by a group of particles, so the terms "particle" and "swarm" respectively indicate individual and population (Abad et al., 2021a; Onwunalu and Durlofsky, 2010). Although the PSO method resembles genetic algorithms in terms of some properties, none of the evaluation operators, cross-over and mutation, are used in the PSO. In the PSO optimization technique, each particle moves around in the optimization problem's domain area based on the influence of its topological neighborhoods, which are physical, social, and queen neighborhoods (Sharma and Onwubolu, 2009).

PSO optimizer defines two vectors, a velocity vector $V_i(t)$ and a position vector $X_i(t)$ for each particle of the population. In each iteration of algorithm, these velocity and position are updated for each particle in the population. The velocity and the position of the particles can be updated based on the following formulas (Eqs. (11) and (12)).

$$X_i\,(t+1) = X_i\,(t) + V_i(t+1) \tag{11}$$

$$V_i\,(t+1) = wV_i\,(t) + c_1 r_1\,(Pb_i\,(t) - X_i\,(t)) + c_2 r_2 (G_b\,(t) - X_i\,(t)) \tag{12}$$

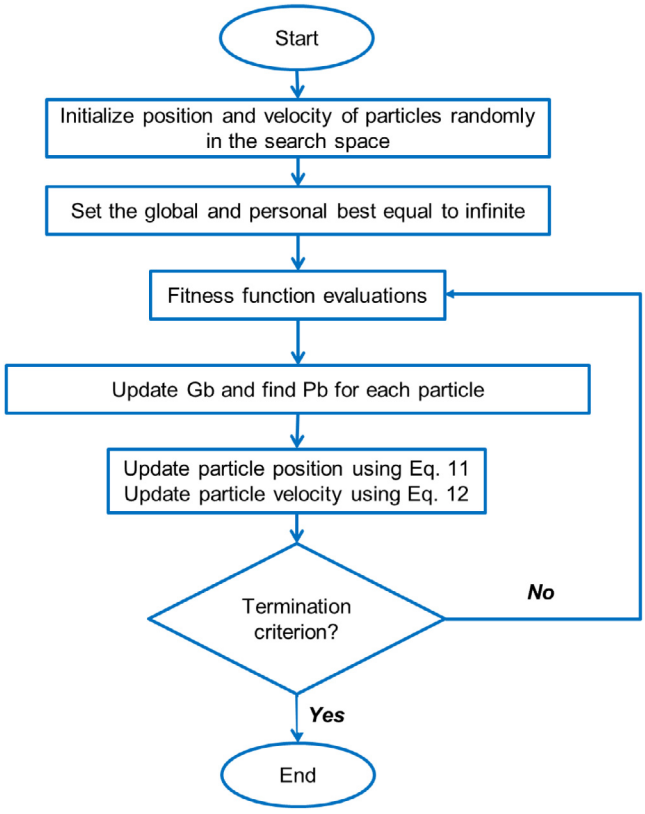

Fig. 2. Illustration of PSO algorithm implementation workflow.

where;

G_b and P_b represent the best global position and the best previous position for the i^{th} particle respectively; w denotes the inertia weight; r stands for random number; and c denotes learning rate (Chen, 2013; Farsi et al., 2021b). The above-mentioned relationship (Eq. (12)) is formed of three components, which are social, inertia, and cognitive components. Inertia component is represented by $wV_i(t)$, which is the previous movements' retention and directs the individual (particle) to its path at tth iteration. The second term, $c_1r_1(Pb_i(t) - X_i(t))$, represents the cognitive component that contains c_1. Movement of the individuals to previous best position is performed by this cognitive component. The third term, $c_2r_2(G_b(t) - X_i(t))$, reparents the social component. This component evaluates the particles efficiency and the swarm trajectory in the domain. Fig. 2 illustrates the flow diagram to use the PSO algorithm.

3. Data analysis

3.1. Data collection

In order to predict PP in this paper, a 2875 dataset related to three wells (S1, S2 and S3) located in the Tabnak gas field in the southeast of Mohr City and north of Fars City, which was discovered in 1978. So far, 44 wells have been drilled in this field, of which 43 are producing gas. The available data is between depths of 3456–3738.6 m (942 data points) in the S1 well, 3462–3764.4 m (1008 data points) in the S2 well, and 3477–3754.5 m (925 data points) in the S3 well. The data collected to predict the PP parameter from log data is related to 12 input variables: the photoelectric index (PEF), rate of penetration (ROP), deep resistivity (ILD), pore compressibility (Cp), weight on bit (WOB), bulk density (RHOB), corrected gamma ray (CGR), laterolog shallow (ILS), uncorrected spectral gamma-ray (SGR), neutron porosity (NPHI), caliper (CALL), hole size (HS), shear-wave velocity (Vs) and compression-wave velocity (Vp). But in order to determine PP, Eaton's method is used. In order to verify PP data, several data points related to data recorded by repeat formation tester (RFT) tools are used. Fig. 3 shows the verification information for some points in the S1, S2 and S3 wells.

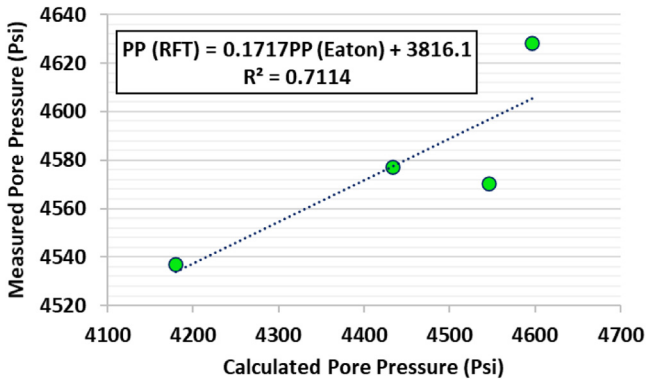

Fig. 3. Pore pressure calculated by Eaton's method datapoint validation by RFT tools.

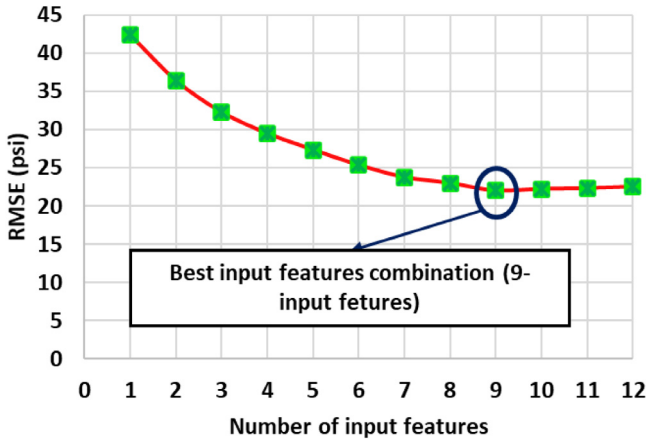

Fig. 4. RMSE vs. number of input variable based on feature selection in wells S1 and S2.

3.2. Feature selection

In order to eliminate unnecessary inputs in this study, we have used the feature selection method. Table 1 shows the notation information for the inputs. Using the feature selection shown in Table 1 and Fig. 4, the PP was predicted. After the results of feature selection are shown, the best number of feature selection is nine input features. The best nine input features are A2, A4, A7, A1, A8, A12, A6, A3 and A11 (ROP, ILD, RHOB, PEF, CGR, vp, WOB, vs and Cp).

After selecting the attribute selection and selecting the 9 attributes, the parameters related to the input variables, including ROP, ILD, RHOB, PEF, CGR, vp, WOB, vs and Cp for three, are shown in Table 2. In order to build single and hybrid artificial intelligence models, data related to two wells (S1 and S2) has been used. Finally, in order to generalize the best algorithm, we used S3-well data to find out if this algorithm gives the same result for other data in this field.

Table 1
Results of feature selection and the determination characters of parameters in wells S1 and S2.

No. of features	Features	RMSE (psi)	Parameters	Characters
1	A12	42.355	PEF	A1
2	A3, A12	36.395	ROP	A2
3	A2, A12, A3	32.259	vs.	A3
4	A12, A3, A2, A6	29.482	ILD	A4
5	A2, A6, A12, A3, A8	27.346	LLS	A5
6	A3, A8, A2, A6, A12, A7	25.338	WOB	A6
7	A7, A12, A6, A2, A3, A8, A1	23.755	RHOB	A7
8	A3, A8, A6, A12, A1, A7, A2, A4	22.974	CGR	A8
9	A2, A4, A7, A1, A8, A12, A6, A3, A11 (Best)	22.001	CALI	A9
10	A12, A6, A3, A7, A2, A11, A8, A4, A1, A5	22.224	HS	A10
11	A4, A12, A2, A8, A7, A4, A6, A5, A3, A1, A9	22.311	Cp	A11
12	A8, A4, A5, A3, A6, A4, A11, A7, A2, A10, A1, A12	22.523	vp	A12

Table 2
Statistical parameters for all input and output variable for wells in wells S1, S2 and S3 after feature selection.

Wells	Variables	The photoelectric index	Rate of penetration	Deep resistivity	Pore compressibility	Weight on bit	Bulk density	Corrected gamma ray	Shear-wave velocity	Compressional-wave velocity	Pore pressure
	Symbol	PEF	ROP	ILD	Cp	WOB	RHOB	CGR	vs	vp	PP
	Units	Barn/cm^3	ft/s	mmho/m	psi-1	klb	g/cm^3	GAPI	km/s	km/s	Psi
Well S1 (942 data point)	Mean	4.13	16.79	847.13	1.19E−06	13.66	2.58	23.45	97.93	53.50	4674.39
	Std. Dev.	0.73	2.28	3567.09	2.85E−07	7.17	0.11	22.18	6.48	3.15	346.31
	Variance	2.29	5.34	1.31	1.32E−07	5.15	2.01	1.06	79.58	47.02	4092.16
	Minimum	6.33	28.24	20000.00	2.23E−06	22.16	2.83	121.40	187.09	67.30	5428.00
	Maximum	−0.27	0.11	4.86	2.48E−01	0.63	−0.91	2.25	3.34	0.57	0.13
	Skewness	0.06	−0.08	22.56	−3.55E−01	−0.35	1.17	5.06	37.28	0.15	−1.11
	Kurtosis	4.31	19.17	1447.61	1.19E−06	17.24	2.88	21.11	396.88	53.38	4872.79
Well S2 (1008 data point)	Mean	1.14	2.65	4802.25	1.56E−06	9.74	0.53	15.85	266.88	3.75	195.56
	Std. Dev.	1.21	1.67	0.45	1.22E−07	4.32	1.20	3.31	62.45	45.72	4498.00
	Variance	7.41	36.67	20012.34	4.07E−06	30.16	3.87	110.20	738.98	82.91	5428.00
	Minimum	0.29	1.17	3.45	−3.21E−01	−0.42	−0.01	1.99	−0.25	2.72	0.54
	Maximum	−0.92	6.44	10.23	6.45E−01	0.15	−0.43	4.84	−1.86	13.75	−0.45
	Skewness	4.90	15.82	1273.94	1.85E−06	16.24	3.49	26.16	283.30	53.82	4759.19
	Kurtosis	1.38	5.65	4571.83	1.65E−07	10.52	0.38	19.89	259.95	1.79	179.59
Well S3 (925 data point)	Mean	4.45	4.32	0.42	1.65E−06	5.34	2.29	3.17	57.55	48.06	4405.47
	Std. Dev.	5.34	27.32	20002.09	2.32E−06	27.15	3.93	124.27	712.52	65.54	5226.00
	Variance	0.44	1.24	3.72	−1.88E−01	0.64	−1.62	2.09	0.47	0.91	0.17
	Minimum	−0.36	6.18	12.15	−6.88E−01	−0.14	1.69	5.30	−1.72	4.03	0.04
	Maximum	4.13	16.79	847.13	1.19E−06	13.66	2.58	23.45	97.93	53.50	4674.39
	Skewness	0.73	2.28	3567.09	2.85E−07	7.17	0.11	22.18	6.48	3.15	346.31
	Kurtosis	2.29	5.34	1.31	1.32E−07	5.15	2.01	1.06	79.58	47.02	4092.16

4. Result and discussion

In order to compare the artificial intelligence algorithms, we used conventional statistical metrics that are the basis of comparison for regression. Among the statistical metrics used to predict PP in this article are the following (Eqs. (13)–(19)).

Relative error (RE) :

$$PD_i = \frac{H_{(Measured)} - H_{(Predicted)}}{H_{(Measured)}} x100 \tag{13}$$

Average relative error (ARE) :

$$ARE = \frac{\sum_{i=1}^{n} PD_i}{n} \tag{14}$$

Absolute average relative error (AARE) :

$$AARE = \frac{\sum_{i=1}^{n} |PD_i|}{n} \tag{15}$$

Coefficient of Determination(R^2) :

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (S_{Predicted\,i} - S_{Measured\,i})^2}{\sum_{i=1}^{N} (S_{Predicted\,i} - \frac{\sum_{I=1}^{n} S_{Measured\,i}}{n})^2} \tag{16}$$

Mean Square Error (MSE) :

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Z_{Measured\,i} - Z_{Predicted\,i})^2 \tag{17}$$

Root Mean Square Error (RMSE) :

$$RMSE = \sqrt{MSE} \tag{18}$$

Standard Deviation (STD) :

$$STD = \sqrt{\frac{\sum_{i=1}^{n} (D_i - Dimean)^2}{n-1}} \tag{19}$$

$$Dimean = \frac{1}{n} \sum_{i=1}^{n} \left(S_{Measured_i} - S_{Predicted_i}\right)$$

The difference between test and validation data is that validation data is entered into the model during training if it has not previously been evaluated, where the first test is provided by this data. This data allows researchers to evaluate useful information up to the model prediction. On the other hand, test data are unlabeled data that are used to perform the final evaluation; in other words, it is predetermined data that is randomly selected for the purpose of final control and output accuracy determination.

Two of the most important metrics used in this paper are RMSE and R^2. In order to find the best algorithm for PP prediction, the results used in Tables 3 to 6 are used. The results in Tables 3 to 6 are the results of statistical tests related to testing, training, validating, and total dataset, respectively. In order to build artificial intelligence algorithms and train them, 1364 data points related to S1 and S2 wells have been used. Furthermore, in order to check the results of these algorithms, 293 data points

Table 3
Prediction accuracy to prediction of PP for six ML and HML models based on six statical metric for training dataset in wells S1 and S2 (1364 data ∼ 70%).

Models Units	ARE (%)	AARE (%)	STD (Psi)	MSE (Psi)	RMSE (Psi)	R^2 –
KNN	0.136	3.522	28.475	8.114E+02	28.4846	0.7903
WKNN	−0.035	3.019	24.361	5.931E+02	24.3541	0.8315
DWKNN	0.013	2.499	20.365	4.134E+02	20.3311	0.8533
KNN-PSO	−0.052	2.013	17.685	3.129E+02	17.6888	0.8964
WKNN-PSO	−0.052	2.305	16.190	2.621E+02	16.1904	0.9016
DWKNN-PSO	−0.022	1.760	12.679	1.605E+02	12.6690	0.9336

Table 4
Prediction accuracy to prediction of PP for six ML and HML models based on six statical metric for testing dataset in wells S1 and S2 (293 data ∼ 15%).

Models Units	ARE (%)	AARE (%)	STD (Psi)	MSE (Psi)	RMSE (Psi)	R^2 –
KNN	−0.214	3.415	25.308	6.423E+02	25.3437	0.8215
WKNN	−0.041	3.203	23.272	5.417E+02	23.2739	0.8741
DWKNN	0.083	2.483	18.308	3.355E+02	18.3160	0.8915
KNN-PSO	−0.133	2.075	16.317	2.671E+02	16.3423	0.9262
WKNN-PSO	−0.133	2.284	14.745	2.178E+02	14.7585	0.9274
DWKNN-PSO	−0.009	1.747	11.257	1.267E+02	11.2570	0.9846

Table 5
Prediction accuracy to prediction of PP for six ML and HML models based on six statical metric for validating dataset in wells S1 and S2 (293 data ∼ 15%).

Models Units	ARE (%)	AARE (%)	STD (Psi)	MSE (Psi)	RMSE (Psi)	R^2 –
KNN	0.144	3.510	26.132	6.837E+02	26.1468	0.8024
WKNN	−0.079	3.023	22.468	5.050E+02	22.4733	0.8533
DWKNN	0.121	2.557	19.017	3.622E+02	19.0314	0.8846
KNN-PSO	0.218	1.989	15.987	2.581E+02	16.0650	0.9137
WKNN-PSO	0.218	2.291	14.918	2.226E+02	14.9183	0.9279
DWKNN-PSO	0.099	1.796	11.745	1.382E+02	11.7571	0.9787

Table 6
Prediction accuracy to prediction of PP for six ML and HML models based on six statical metric for total dataset in wells S1 and S2 (1950 data = 100%).

Models Units	ARE (%)	AARE (%)	STD (Psi)	MSE (Psi)	RMSE (Psi)	R^2 –
KNN	0.136	3.522	28.475	8.114E+02	28.4846	0.8047
WKNN	−0.035	3.019	24.361	5.931E+02	24.3541	0.8530
DWKNN	0.013	2.499	20.365	4.134E+02	20.3311	0.8765
KNN-PSO	−0.052	2.013	17.685	3.129E+02	17.6888	0.9121
WKNN-PSO	−0.052	2.305	16.190	2.621E+02	16.1904	0.9190
DWKNN-PSO	−0.024	1.762	12.687	1.607E+02	12.6773	0.9656

Table 7
Generalization of DWKNN-PSO based on all dataset of well S3 for determination of accuracy to prediction of PP based on training data (70% dataset from wells S1 and S2).

Models Units	ARE (%)	AARE (%)	STD (Psi)	MSE (Psi)	RMSE (Psi)	R^2 –
DWKNN-PSO	−0.019	1.001	9.759	9.515E+01	9.7545	0.9756

known as testing data are used, and finally, in order to validate these algorithms, 293 data points are used as validation data.

The results of each of the algorithms in Table 3 for the training dataset show the high-performance accuracy of algorithm DWKNN-PSO compared to other algorithms. As indicated in the results, the best result for DWKNN-PSO is presented in (Table 3), these values are AARE (KNN) = 1.76 psi and R^2 (KNN) = 0.9336.

The results of each algorithm in Table 3 for the training dataset show that algorithm DWKNN-PSO outperforms other algorithms in terms of accuracy.

As shown in the results, the best DWKNN-PSO result is presented in (Table 3), with values of AARE (KNN) = 1.76 psi and R2 (KNN) = 0.9336.

Table 5 displays the validation dataset results, and this table demonstrates the high-performance accuracy of the DWKNN-PSO algorithm, with RMSE, ARE, AARE, and STD values of 11.757, 0.099, 1.796, and 11.745, respectively. When the DWKNN-PSO algorithm's accuracy is compared to the other five algorithms when tested on the validation subset, it is clear that this algorithm outperforms the others in terms of PP prediction accuracy. For instance, the STD value achieved by the DWKNN-PSO is 11.745, whereas the other five algorithms evaluated, KNN, WKNN, DWKNN, KNN-PSO, and WKNN-PSO present the STD values of 26.132, 22.468, 19.017, 15.987,14.918, respectively.

Finally, Table 6 shows the results for six ML and HML models based on six statistical metrics for the total dataset. This table also shows the high-performance accuracy of this article for the DWKNN-PSO algorithm. The final results for total data is for each algorithm are: AARE (KNN) = 1.796 psi and R^2 (KNN) = 0.9787, AARE (WKNN) = 1.796 psi and R^2 (WKNN) = 0.9787, AARE (DWKNN) = 1.796 psi and R^2 (DWKNN) = 0.9787, AARE (KNN-PSO) = 1.796 psi and R^2 (KNN-PSO) = 0.9787, AARE (WKNN-PSO) = 1.796 psi and R^2 (WKNN-PSO) = 0.9787, AARE (DWKNN-PSO) = 1.762 psi and R^2 (DWKNN-PSO) = 0.9665.

Furthermore, comparing the values of the other accuracy metrics used, RMSE, STD, and ARE, DWKNN-PSO produces the most accurate PP predictions. For example, the STD, RMSE, and ARE values achieved by the DWKNN-PSO in PP prediction when tested on the entire dataset are 12.687, 12.6773, and −0.024, respectively, which are lower than those achieved by the other five algorithms tested. Fig. 5 shows a cross diagram for predicting pp versus the verified value for all datasets for the S1 and S2 wells. From this figure, several conclusions can be drawn, and it can show a visual image of the algorithms. The DWKNN-PSO algorithm, like the entire visual form, has a much higher performance accuracy than other hybrid or even single algorithms. Further, this diagram shows the coefficient of determination of each algorithm for the whole dataset. From the information about this parameter, it can be concluded that the performance accuracy of the HML is greater than other algorithms. Algorithms can be classified in terms of this statistical parameter as follows: KNN < WKNN < DWKNN < KNN-PSO < WKNN-PSO < DWKNN-PSO.

Fig. 6 shows the error rate and R^2 values for the different algorithms. It can be noted that the low RMSE error rate is inversely proportional to the R^2 value (for all datasets of wells S1 and S2). In other words, if the error obtained by each algorithm is reduced, a better fit in the results can be obtained. Based on this diagram, the performance accuracy (RMSE) of DWKNN-PSO is higher than that of other algorithms, and the fit of the diagram is also higher. In other words, the results of this diagram can be shown as follows:

R^2: R^2 (DWKNN-PSO) > R^2 (WKNN-PSO) > R^2 (KNN-PSO) > R^2 (DWKNN) > R^2 (WKNN) > R^2 (KNN)

RMSE: RMSE (DWKNN-PSO) < RMSE (WKNN-PSO) < RMSE (KNN-PSO) < RMSE (DWKNN)< RMSE (WKNN) < RMSE (KNN)

Fig. 6 also confirms the error between the different algorithms and compares the AAPD and STD errors using the Pie diagram (for all datasets of wells S1 and S2). This diagram shows that the DWKNN-PSO algorithm has the highest performance accuracy in terms of STD = 12.687 psi and AARE = 1.762%, while the KNN algorithm has the lowest performance accuracy. This chart can visually and quickly present the results (see Fig. 7).

5. Generalize DWKNN-PSO model for predicted PP

The results shown in the previous section discuss training, testing, and validation for three ML (KNN, WKK, and DWKNN) and three HML (KNN-PSO, WKNN-PSO, and DWKNN-PSO) based

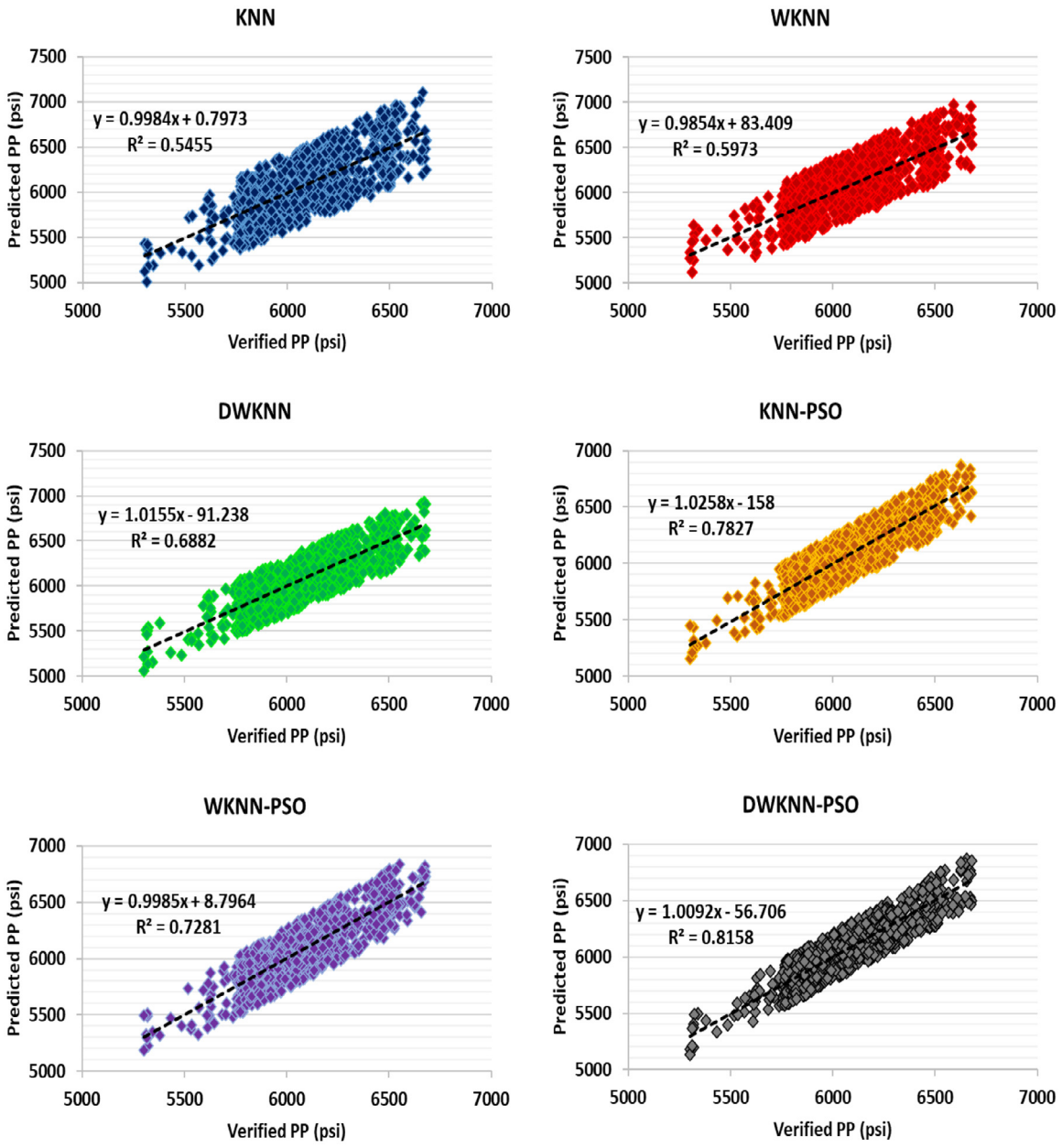

Fig. 5. Illustration of predicted and verify PP for three ML (KNN, WKK and DWKNN) and three HML (KNN-PSO, WKNN-PSO and DWKNN-PSO) based on all of dataset in well S1 and well S2.

on the datasets from the S1 and S2 wells. After reviewing the results presented in this model, it can be observed that the best artificial intelligence algorithm to predict this important parameter is DWKNN-PSO. In this section, we tried to investigate if the best algorithm for predicting PP could offer the same level of performance accuracy as other wells in the studied field. The results presented in Table 7 confirm the performance accuracy of the DWKNN-PSO algorithm in S3 well with the results presented in the training, testing, and validation section in the S2 and S1 wells.

Fig. 8 also confirms the performance accuracy of the new algorithm presented in predicting this critical parameter (PP). The findings of this study indicate that researchers can apply this algorithm to predict other important reservoir/geological parameters.

6. Recommendations for future research works

Evaluation of the effect of the inclusion of other drilling parameters such as standpipe pressure and mud flow rate as input parameters along with well logging to predict PP can be further investigated. According to the current findings, adding more related input parameters could provide models with higher prediction efficiencies. Involving other optimizers, such as genetic algorithms and firefly algorithms, in the development of a high-performance hybrid predictive model for PP prediction can also be considered in future research work (Ahmadi et al., 2020; Choubin et al., 2019; Kalbasi et al., 2021; Mosavi et al., 2021; Qasem et al., 2019). The application of the proposed method should be investigated in a wide range of applications, e.g., various energy, ecological and natural research applications (Ahmadi et al., 2020; Band et al., 2020a,b; Emadi et al., 2020; Lei

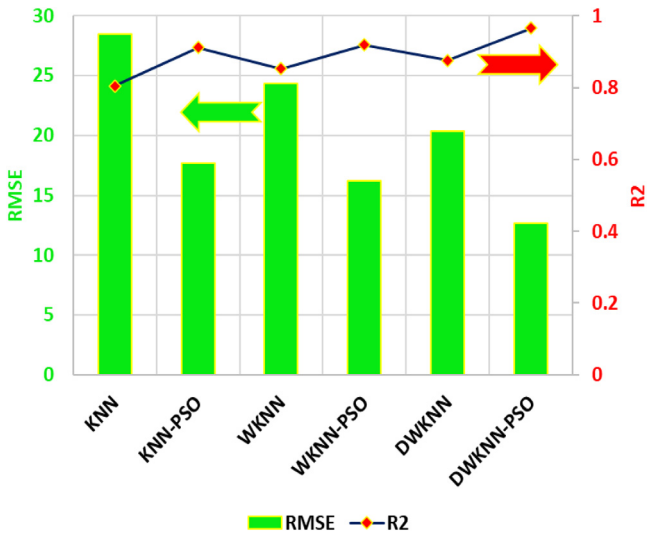

Fig. 6. Illustration of RMSE & R^2 for predicted of PP for three ML (KNN, WKK and DWKNN) and three HML (KNN-PSO, WKNN-PSO and DWKNN-PSO) based on all of dataset in well S1 and well S2.

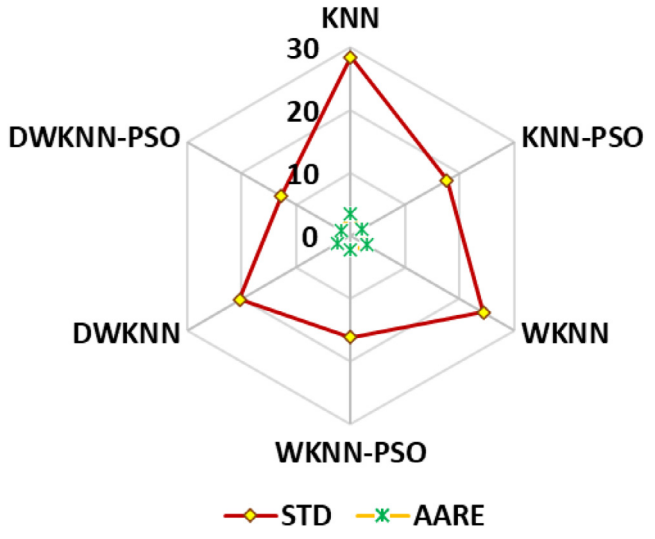

Fig. 7. Illustration of AARE for predicted of PP for three ML (KNN, WKK and DWKNN) and three HML (KNN-PSO, WKNN-PSO and DWKNN-PSO) based on all of dataset in well S1 and well S2.

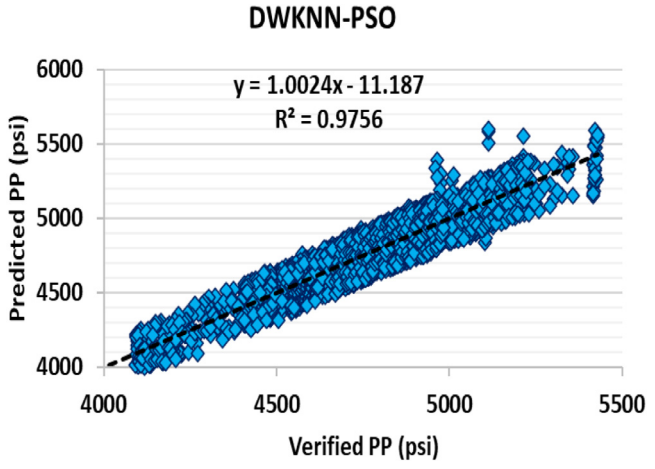

Fig. 8. Illustration of predicted and verified PP to generalized the beast HML algorithm (DWKNN-PSO) based on well S3 data records.

et al., 2020; Shamshirband et al., 2020; Taherei Ghazvinei et al., 2018). From computational fluid and hydrological modeling to environmental simulation for instance (Ghalandari et al., 2019c; Mahmoudi et al., 2021; Rezakazemi et al., 2019; Seifi et al., 2020) the proposed methodology can be effective. For the future research the comparative analysis with other machine learning methods, e.g., Asadi et al. (2019), Ghalandari et al. (2019b), Joloudari et al. (2020), Mosavi et al. (2020), Mosavi and Safaei-Farouji (2021), Sadeghzadeh et al. (2020) and Shabani et al. (2020) would be essential to bring an insight into the true potential of the proposed method (Ghalandari et al., 2019a; Nabipour et al., 2020b). To improve the accuracy and the performance of the proposed method further deep learning, ensemble and hybrid methods for instance, those suggest in Band et al. (2020b), Dehghani et al. (2020), Mosavi et al. (2020), Mousavi et al. (2021), Nabipour et al. (2020a), Nourani et al. (2022) and Shamsirband and Mehri Khansari (2021) can come to the consideration.

7. Conclusions

This study aims to predict the PP applying six novel algorithms, including three ML (KNN, WKK, and DWKNN) and three HML (KNN-PSO, WKNN-PSO, and DWKNN-PSO) based on log and drilling data. To the best of the authors knowledge, these algorithms have not been suggested to predict this parameter so far. Three ML algorithms and their hybrid forms were developed using based on 1950 (70%) data related to S1 and S2 wells for building algorithms (algorithm training), 293 data (15%) for testing, and 293 (15%) for validation. After reviewing the input data with the feature selection method, the most influential nine input features are: rate of penetration (ROP), deep resistivity (ILD), density (RHOB), the photoelectric index (PEF), corrected gamma ray (CGR), compression-wave velocity (Vp), weight on bit (WOB), shear-wave velocity (Vs) and pore compressibility (Cp) and finally, using these best inputs, AI models are developed. In this paper, the ROP and WOB were used, where the study performed demonstrates that these two parameters are also effective for predicting PP. The KNN based algorithm can be used for evaluating datasets with noisy or incomplete data samples. The main concerns with KNN based algorithms are the negative effects of K values and neighbors' weights on the accuracy of predicted target values. As a result, the PSO optimizer was employed in this study to estimate optimal weights and K values to enhance the model's prediction accuracy. When the results of three ML algorithms and three HML algorithms were compared, it was discovered that HML algorithms outperformed ML algorithms, where the DWKNN-PSO was found to be the best algorithm in predicting PP among the six algorithms tested (for total dataset of wells S1 and S2: R2 = 0.9656 and RMSE = 12.6773 psi). Finally, in order to test the generalizability of the best-performing algorithm for other wells in the field under study, we tested the DWKNN-PSO on the unseen data from the S3 well, the results of which confirm the performance accuracy of this algorithm (R2 = 0.9765 and RMSE = 9.7545 psi for the total dataset of wells S3). Considering the promising accuracy presented by the algorithms used in this study, the developed models can be considered to predict PP in the studied field or be applied to other fields.

Nomenclature	
ANFIS	Adaptive neuro fuzzy inference system
ANN	Artificial neural network
DE	Drilling efficiency
DWKNN	Distance weight K-nearest neighbor
ELM	Extreme learning machine
FL	Fuzzy logic
FN	Functional networks
FP	Fracture pressure
GR	Gamma-ray
HML	Hybrid machine learning
KNN	K-nearest neighbor
LSSVM	Least squares support-vector machine
ML	Machine learning
MLP	Multilayer perceptron
MLP-ANN	Multilayer perceptron neural network
MSE	Mechanical specific energy
NPHI	Neutron porosity
P_{eff}	Effective pressure
P_{over}	Overburden pressure
PP	Pore pressure
PSO	Particle swarm optimization
PSO	Particle Swarm Optimization
RBF	Radial basis function
SC	Slowness of compression
SVM	Support vector machine
SVM	Support vector machine
Vs	Shear wave velocity
WKNN	Weight K-nearest neighbor
C_P	Predicted value of output variable
C_t	Values of output variable of the nearest neighbor t
D_i	Calculated Euclidean distance
$V_i(t)$	Velocity vector
W_i	Dependent variable weight
$X_i(t)$	Position vector
x_i	Test sample value
y_i	The value of ith training sample
ω	Biot coefficient
c	Learning rate
r	Random number

CRediT authorship contribution statement

Farshad Jafarizadeh: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Meysam Rajabi:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Somayeh Tabasi:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Reza Seyedkamali:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Shadfar Davoodi:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Hamzeh Ghorbani:** Conceptualization, Supervision, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Mehdi Ahmadi Alvar:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Ahmed E. Radwan:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Mako Csaba:** Conceptualization, Data curation, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Since the data used in the study are confidential, the authors cannot share them publicly.

Acknowledgment

This research was supported by the Tomsk Polytechnic University development program. The corresponding author, Ahmed E. Radwan, thanks the Jagillionian University for funding the Article Processing Charge (APC).

References

Abad, A.R.B., Ghorbani, H., Mohamadian, N., Davoodi, S., Mehrad, M., Aghdam, S.K.-y., Nasriani, H.R., 2022. Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields. Fuel 308, 121872. http://dx.doi.org/10.1016/j.fuel.2021.121872.

Abad, A.R.B., Mousavi, S., Mohamadian, N., Wood, D.A., Ghorbani, H., Davoodi, S., Alvar, M.A., Shahbazi, K., 2021a. Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate reservoirs. J. Nat. Gas Sci. Eng. 95, 104210.

Abad, A.R.B., Tehrani, P.S., Naveshki, M., Ghorbani, H., Mohamadian, N., Davoodi, S., Aghdam, S.K.-y., Moghadasi, J., Saberi, H., 2021b. Predicting oil flow rate through orifice plate with robust machine learning algorithms. Flow Meas. Instrum. 102047. http://dx.doi.org/10.1016/j.flowmeasinst.2021.102047.

Abdali, M.R., Mohamadian, N., Ghorbani, H., Wood, D.A., 2021. Petroleum well blowouts as a threat to drilling operation and wellbore sustainability: causes, prevention, safety and emergency response. J. Constru. Mater.| Special Issue on Sustainable Petroleum Engineering ISSN 2652, 3752.

Abdelghany, W.K., Radwan, A.E., Elkhawaga, M.A., Wood, D.A., Sen, S., Kassem, A.A., 2021. Geomechanical modeling using the depth-of-damage approach to achieve successful underbalanced drilling in the Gulf of Suez rift basin. J. Pet. Sci. Eng. 202, 108311.

Abidin, M.H., 2014. Pore pressure estimation using artificial neural network. http://utpedia.utp.edu.my/id/eprint/14317.

Ahmadi, M.H., Baghban, A., Sadeghzadeh, M., Zamen, M., Mosavi, A., Shamshirband, S., Kumar, R., Mohammadi-Khanaposhtani, M., 2020. Evaluation of electrical efficiency of photovoltaic thermal solar collector. Eng. Appl. Comput. Fluid Mech. 14, 545–565. http://dx.doi.org/10.1080/19942060.2020.1734094.

Asadi, E., Isazadeh, M., Samadianfard, S., Ramli, M.F., Mosavi, A., Nabipour, N., Shamshirband, S., Hajnal, E., Chau, K.-W., 2019. Groundwater quality assessment for sustainable drinking and irrigation. Sustainability 12, 177. http://dx.doi.org/10.3390/su12010177.

Bahrami, B., Sadatshojaie, A., Wood, D.A., 2020. Assessing wellbore stability with a modified lade failure criterion. J. Energy Resour. Technol. 142, 083004. http://dx.doi.org/10.1115/1.4046387.

Band, S.S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakrabortty, R., Melesse, A.M., Mosavi, A., 2020a. Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. Remote Sens. 12 (3568), http://dx.doi.org/10.3390/rs12213568.

Band, S.S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakrabortty, R., Shokri, M., Mosavi, A., 2020b. Novel ensemble approach of deep learning neural network (DLNN) model and particle swarm optimization (PSO) algorithm for prediction of gully erosion susceptibility. Sensors 20, 5609. http://dx.doi.org/10.3390/s20195609.

Baouche, R., Sen, S., Ganguli, S.S., 2020. Pore pressure and in-situ stress magnitudes in the Bhiret Hammou hydrocarbon field, Berkine Basin, Algeria. J. Afr. Earth Sci. 171, 103945. http://dx.doi.org/10.1016/j.jafrearsci.2020.103945.

Barjouei, H.S., Ghorbani, H., Mohamadian, N., Wood, D.A., Davoodi, S., Moghadasi, J., Saberi, H., 2021. Prediction performance advantages of deep machine learning algorithms for two-phase flow rates through wellhead chokes. J. Pet. Explor. Prod. 11, 1233–1261, https://link.springer.com/article/10.1007/s13202-021-01087-4.

Biot, M.A., Willis, D.G., 1957. The elastic coefficients of the theory of consolidation. http://dx.doi.org/10.1115/1.4011606.

Brainard, R., 2006. A process used in evaluation of managed-pressure drilling candidates and probabilistic cost–benefit analysis. In: Offshore Technology Conference. Offshore Technology Conference. http://dx.doi.org/10.4043/18375-MS.

Brajard, J., Carrassi, A., Bocquet, M., Bertino, L., 2020. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the lorenz 96 model. J. Comput. Sci. 44, 101171. http://dx.doi.org/10.1016/j.jocs.2020.101171.

Chen, M.-Y., 2013. A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. Inform. Sci. 220, 180–195. http://dx.doi.org/10.1016/j.ins.2011.09.013.

Choubin, B., Mosavi, A., Alamdarloo, E.H., Hosseini, F.S., Shamshirband, S., Dashtekian, K., Ghamisi, P., 2019. Earth fissure hazard prediction using machine learning models. Environ. Res. 179, 108770. http://dx.doi.org/10.1016/j.envres.2019.108770.

Dawson, E.M., Roth, W.H., Nesarajah, S., Bureau, G., Davis, C.A., 2020. A practice oriented pore-pressure generation model. In: FLAC and Numerical Modeling in Geomechanics. CRC Press, pp. 47–54.

Dehghani, M., Salehi, S., Mosavi, A., Nabipour, N., Shamshirband, S., Ghamisi, P., 2020. Spatial analysis of seasonal precipitation over Iran: Co-variation with climate indices. ISPRS Int. J. Geo-Inf. 9 (73), http://dx.doi.org/10.3390/ijgi9020073.

Du, W., Slaný, M., Wang, X., Chen, G., Zhang, J., 2020. The inhibition property and mechanism of a novel low molecular weight zwitterionic copolymer for improving wellbore stability. Polymers 12 (708), http://dx.doi.org/10.3390/polym12030708.

Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory. In: MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science. Ieee, pp. 39–43. http://dx.doi.org/10.1109/MHS.1995.494215.

Elkatatny, S., Ahmed, A., Abughaban, M., Patil, S., 2020. Deep illustration for loss of circulation while drilling. Arab. J. Sci. Eng. 45, 483–499, https://link.springer.com/article/10.1007/s13369-019-04315-6.

Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., Scholten, T., 2020. Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. Remote Sens. 12 (2234), http://dx.doi.org/10.3390/rs12142234.

Farsi, M., Barjouei, H.S., Wood, D.A., Ghorbani, H., Mohamadian, N., Davoodi, S., Nasriani, H.R., Alvar, M.A., 2021a. Prediction of oil flow rate through orifice flow meters: Optimized machine-learning techniques. Measurement 174, 108943. http://dx.doi.org/10.1016/j.measurement.2020.108943.

Farsi, M., Mohamadian, N., Ghorbani, H., Wood, D.A., Davoodi, S., Moghadasi, J., Alvar, M.A., 2021b. Predicting formation pore-pressure from well-log data with hybrid machine-learning optimization algorithms. Nat. Resour. Res. 1–27, https://link.springer.com/article/10.1007/s11053-021-09852-2.

Ghalandari, M., Shamshirband, S., Mosavi, A., Chau, K.-w., 2019a. Flutter speed estimation using presented differential quadrature method formulation. Eng. Appl. Comput. Fluid Mech. 13, 804–810. http://dx.doi.org/10.1080/19942060.2019.1627676.

Ghalandari, M., Ziamolki, A., Mosavi, A., Shamshirband, S., Chau, K.-W., 2019b. Aeromechanical optimization of first row compressor test stand blades. http://dx.doi.org/10.20944/preprints201905.0049.v1.

Ghalandari, M., Ziamolki, A., Mosavi, A., Shamshirband, S., Chau, K.-W., Bornassi, S., 2019c. Aeromechanical optimization of first row compressor test stand blades using a hybrid machine learning model of genetic algorithm, artificial neural networks and design of experiments. Eng. Appl. Comput. Fluid Mech. 13, 892–904. http://dx.doi.org/10.1080/19942060.2019.1649196.

Ghorbani, H., Moghadasi, J., Wood, D.A., 2017. Prediction of gas flow rates from gas condensate reservoirs through wellhead chokes using a firefly optimization algorithm. J. Nat. Gas Sci. Eng. 45, 256–271. http://dx.doi.org/10.1016/j.jngse.2017.04.034.

Ghorbani, H., Wood, D.A., Choubineh, A., Mohamadian, N., Tatar, A., Farhangian, H., Nikooey, A., 2020. Performance comparison of bubble point pressure from oil PVT data: Several neurocomputing techniques compared. Exp. Comput. Multiph. Flow 2, 225–246, https://link.springer.com/article/10.1007/s42757-019-0047-5.

Ghorbani, H., Wood, D.A., Moghadasi, J., Choubineh, A., Abdizadeh, P., Mohamadian, N., 2019. Predicting liquid flow-rate performance through wellhead chokes with genetic and solver optimizers: an oil field case study. J. Pet. Explor. Prod. Technol. 9, 1355–1373, https://link.springer.com/article/10.1007/s13202-018-0532-6.

Hazbeh, O., Aghdam, S.K.-y., Ghorbani, H., Mohamadian, N., Alvar, M.A., Moghadasi, J., 2021a. Comparison of accuracy and computational performance between the machine learning algorithms for rate of penetration in directional drilling well. Pet. Res. http://dx.doi.org/10.1016/j.ptlrs.2021.02.004.

Hazbeh, O., Ahmadi Alvar, M., Aghdam, K.-y., Ghorbani, H., Mohamadian, N., Moghadasi, J., 2021b. Hybrid computing models to predict oil formation volume factor using multilayer perceptron algorithm. J. Pet. Min. Eng. 1, 4–27. http://dx.doi.org/10.21608/JPME.2021.52149.1062.

Hou, P., Gao, F., Gao, Y., Yang, Y., Cai, C., 2017. Effect of pore pressure distribution on fracture behavior of sandstone in nitrogen fracturing. Energy Explor. Exploit. 35, 609–626. http://dx.doi.org/10.1177/0144598717709666.

Hu, L., Deng, J., Zhu, H., Lin, H., Chen, Z., Deng, F., Yan, C., 2013. A new pore pressure prediction method-back propagation artificial neural network. Electron. J. Geotech. Eng. 18, 4093–4107, http://www.ejge.com/2013/Ppr2013.371mlr.pdf.

Huang, M., Lin, R., Huang, S., Xing, T., 2017. A novel approach for precipitation forecast via improved K-nearest neighbor algorithm. Adv. Eng. Inform. 33, 89–95. http://dx.doi.org/10.1016/j.aei.2017.05.003.

Joloudari, J.H., Hassannataj Joloudari, E., Saadatfar, H., Ghasemigol, M., Razavi, S.M., Mosavi, A., Nabipour, N., Shamshirband, S., Nadai, L., 2020. Coronary artery disease diagnosis; ranking the significant features using a random trees model. Int. J. Environ. Res. Public Health 17, 731. http://dx.doi.org/10.3390/ijerph17030731.

Kalbasi, R., Jahangiri, M., Mosavi, A., Dehshiri, S.J.H., Dehshiri, S.S.H., Ebrahimi, S., Etezadi, Z.A.-S., Karimipour, A., 2021. Finding the best station in Belgium to use residential-scale solar heating, one-year dynamic simulation with considering all system losses: economic analysis of using ETSW. Sustain. Energy Technol. Assess. 45, 101097. http://dx.doi.org/10.1016/j.seta.2021.101097.

Keshavarzi, R., Jahanbakhshi, R., 2013. Real-time prediction of pore pressure gradient through an artificial intelligence approach: a case study from one of middle east oil fields. Eur. J. Environ. Civ. Eng. 17, 675–686. http://dx.doi.org/10.1080/19648189.2013.811614.

Kiss, A., Fruhwirth, R.K., Pongratz, R., Maier, R., Hofstätter, H., 2018. Formation breakdown pressure prediction with artificial neural networks. In: SPE International Hydraulic Fracturing Technology Conference and Exhibition. OnePetro, https://onepetro.org/ARMAUSRMS/proceedings-abstract/ARMA18/All-ARMA18/ARMA-2018-1098/124075.

Kombo, O.H., Kumaran, S., Sheikh, Y.H., Bovim, A., Jayavel, K., 2020. Long-term groundwater level prediction model based on hybrid KNN-RF technique. Hydrology 7 (59), http://dx.doi.org/10.3390/hydrology7030059.

Kremieniewski, M., 2020. Ultra-lightweight cement slurry to seal wellbore of poor wellbore stability. Energies 13 (3124), http://dx.doi.org/10.3390/en13123124.

Lei, X., Chen, W., Avand, M., Janizadeh, S., Kariminejad, N., Shahabi, H., Costache, R., Shahabi, H., Shirzadi, A., Mosavi, A., 2020. GIS-based machine learning algorithms for gully erosion susceptibility mapping in a semi-arid region of Iran. Remote Sens. 12 (2478), http://dx.doi.org/10.3390/rs12152478.

Mahmoudi, M.R., Heydari, M.H., Qasem, S.N., Mosavi, A., Band, S.S., 2021. Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries. Alex. Eng. J. 60, 457–464. http://dx.doi.org/10.1016/j.aej.2020.09.013.

Mosavi, A., Faghan, Y., Ghamisi, P., Duan, P., Ardabili, S.F., Salwana, E., Band, S.S., 2020. Comprehensive review of deep reinforcement learning methods and applications in economics. Mathematics 8, 1640. http://dx.doi.org/10.3390/math8101640.

Mosavi, A., Safaei-Farouji, M., 2021. Oil family typing using a hybrid model of self-organizing map and artificial neural network. http://dx.doi.org/10.2139/ssrn.3991002, Available at SSRN 3991002.

Mosavi, A., Sajedi Hosseini, F., Choubin, B., Goodarzi, M., Dineva, A.A., Rafiei Sardooi, E., 2021. Ensemble boosting and bagging based machine learning models for groundwater potential prediction. Water Resour. Manage. 35, 23–37. http://dx.doi.org/10.1007/s11269-020-02704-3.

Mousavi, S.M., Ghasemi, M., Dehghan Manshadi, M., Mosavi, A., 2021. Deep learning for wave energy converter modeling using long short-term memory. Mathematics 9 (871), http://dx.doi.org/10.3390/math9080871.

Nabipour, N., Mosavi, A., Hajnal, E., Nadai, L., Shamshirband, S., Chau, K.-W., 2020b. Modeling climate change impact on wind power resources using adaptive neuro-fuzzy inference system. Eng. Appl. Comput. Fluid Mech. 14, 491–506. http://dx.doi.org/10.1080/19942060.2020.1722241.

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E., 2020a. Deep learning for stock market prediction. Entropy 22, 840. http://dx.doi.org/10.3390/e22080840.

Naveshki, M., Naghiei, A., Soltani Tehrani, P., Ahmadi Alvar, M., Ghorbani, H., Mohamadian, N., Moghadasi, J., 2021. Prediction of bubble point pressure using new hybrid computationail intelligence models. J. Chem. Pet. Eng. http://dx.doi.org/10.22059/JCHPE.2021.314719.1341.

Navot, A., Shpigelman, L., Tishby, N., Vaadia, E., 2005. Nearest neighbor based feature selection for regression and its application to neural activity. In: Advances in Neural Information Processing Systems 18. pp. 996–1002.

Nguyen, D., Ouala, S., Drumetz, L., Fablet, R., 2019. Em-like learning chaotic dynamics from noisy and partial observations. arXiv preprint arXiv:1903.10335. https://arxiv.org/abs/1903.10335.

Nour, A., AlBinHassan, N., 2013. Seismic attributes and advanced computer algorithm method to predict formation pore pressure: Paleozoic sediments of northwest saudi arabia. Eur. Assoc. Geosci. Eng. cp–350. http://dx.doi.org/10.3997/2214-4609-pdb.350.iptc16680.

Nourani, M., Alali, N., Samadianfard, S., Band, S.S., Chau, K.-w., Shu, C.-M., 2022. Comparison of machine learning techniques for predicting porosity of chalk. J. Pet. Sci. Eng. 209, 109853. http://dx.doi.org/10.1016/j.petrol.2021.109853.

Onwunalu, J.E., Durlofsky, L.J., 2010. Application of a particle swarm optimization algorithm for determining optimum well location and type. Comput. Geosci. 14, 183–198, https://link.springer.com/article/10.1007/s10596-009-9142-1.

Qasem, S.N., Samadianfard, S., Sadri Nahand, H., Mosavi, A., Shamshirband, S., Chau, K.-w., 2019. Estimating daily dew point temperature using machine learning algorithms. Water 11, 582. http://dx.doi.org/10.1016/j.molliq.2018.11.017.

Radwan, A.E., 2021. Modeling pore pressure and fracture pressure using integrated well logging, drilling based interpretations and reservoir data in the Giant El Morgan oil Field, Gulf of Suez, Egypt. J. Afr. Earth Sci. 104165. http://dx.doi.org/10.1016/j.jafrearsci.2021.104165.

Radwan, A.E., Abudeif, A.M., Attia, M.M., Elkhawaga, M.A., Abdelghany, W.K., Kasem, A.A., 2020. Geopressure evaluation using integrated basin modelling, well-logging and reservoir data analysis in the northern part of the Badri oil field, Gulf of Suez, Egypt. J. Afr. Earth Sci. 162, 103743.

Radwan, A.E., Abudeif, A.M., Attia, M.M., Mohammed, M.A., 2019. Pore and fracture pressure modeling using direct and indirect methods in Badri Field, Gulf of Suez, Egypt. J. Afr. Earth Sci. 156, 133–143. http://dx.doi.org/10.1016/j.jafrearsci.2019.04.015.

Radwan, A., Sen, S., 2021a. Stress path analysis for characterization of in situ stress state and effect of reservoir depletion on present-day stress magnitudes: Reservoir geomechanical modeling in the Gulf of Suez Rift Basin, Egypt. Nat. Resour. Res. 30, 463–478. http://dx.doi.org/10.1007/s11053-020-09731-2.

Radwan, A.E., Sen, S., 2021b. Characterization of in-situ stresses and its implications for production and reservoir stability in the depleted El Morgan hydrocarbon field, Gulf of Suez Rift Basin, Egypt. J. Struct. Geol. 104355. http://dx.doi.org/10.1016/j.jsg.2021.104355.

Radwan, A.E., Wood, D.A., Radwan, A.A., 2022. Machine learning and data-driven prediction of pore pressure from geophysical logs: A case study for the Mangahewa gas field, New Zealand. J. Rock Mech. Rock Eng..

Rajabi, M., Beheshtian, S., Davoodi, S., Ghorbani, H., Mohamadian, N., Radwan, A.E., Alvar, M.A., 2021. Novel hybrid machine learning optimizer algorithms to prediction of fracture density by petrophysical data. J. Pet. Explor. Prod. Technol. 1–23, https://link.springer.com/article/10.1007/s13202-021-01321-z.

Ranaee, E., Ghorbani, H., Keshavarzian, S., Abarghoei, P.G., Riva, M., Inzoli, F., Guadagnini, A., 2021. Analysis of the performance of a crude-oil desalting system based on historical data. Fuel 291, 120046. http://dx.doi.org/10.1016/j.fuel.2020.120046.

Rashidi, S., Mehrad, M., Ghorbani, H., Wood, D.A., Mohamadian, N., Moghadasi, J., Davoodi, S., 2021. Determination of bubble point pressure & oil formation volume factor of crude oils applying multiple hidden layers extreme learning machine algorithms. J. Pet. Sci. Eng. 202, 108425. http://dx.doi.org/10.1016/j.petrol.2021.108425.

Rashidi, S., Mohamadian, N., Ghorbani, H., Wood, D.A., Shahbazi, K., Alvar, M.A., 2020. Shear modulus prediction of embedded pressurized salt layers and pinpointing zones at risk of casing collapse in oil and gas wells. J. Appl. Geophys. 183, 104205. http://dx.doi.org/10.1016/j.jappgeo.2020.104205.

Rezakazemi, M., Mosavi, A., Shirazian, S., 2019. ANFIS pattern for molecular membranes separation optimization. J. Molecular Liquids 274, 470–476. http://dx.doi.org/10.1016/j.molliq.2018.11.017.

Sadeghzadeh, M., Maddah, H., Ahmadi, M.H., Khadang, A., Ghazvini, M., Mosavi, A., Nabipour, N., 2020. Prediction of thermo-physical properties of TiO2-Al2O3/water nanoparticles by using artificial neural network. Nanomaterials 10, 697. http://dx.doi.org/10.3390/nano10040697.

Seifi, A., Ehteram, M., Singh, V.P., Mosavi, A., 2020. Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and ANN. Sustainability 12, 4023. http://dx.doi.org/10.3390/su12104023.

Shabani, S., Samadianfard, S., Sattari, M.T., Mosavi, A., Shamshirband, S., Kmet, T., Várkonyi-Kóczy, A.R., 2020. Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis. Atmosphere 11, 66. http://dx.doi.org/10.3390/atmos11010066.

Shamshirband, S., Mosavi, A., Rabczuk, T., Nabipour, N., Chau, K.-w., 2020. Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines. Eng. Appl. Comput. Fluid Mech. 14, 805–817. http://dx.doi.org/10.1080/19942060.2020.1773932.

Shamsirband, S., Mehri Khansari, N., 2021. Micro-mechanical damage diagnosis methodologies based on machine learning and deep learning models. J. Zhejiang Univ.-Sci. A 22, 585–608, https://link.springer.com/article/10.1631/jzus.A2000408.

Sharma, A., Onwubolu, G., 2009. Hybrid particle swarm optimization and GMDH system. In: Hybrid Self-Organizing Modeling Systems. Springer, pp. 193–231.

Taherei Ghazvinei, P., Hassanpour Darvishi, H., Mosavi, A., Yusof, K.b.W., Alizamir, M., Shamshirband, S., Chau, K.-w., 2018. Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. Eng. Appl. Comput. Fluid Mech. 12, 738–749. http://dx.doi.org/10.1080/19942060.2018.1526119.

Terzaghi, K., 1943. Liner-plate tunnels on the Chicago (Il) subway. Trans. Am. Soc. Civ. Eng. 108, 970–1007. http://dx.doi.org/10.1061/TACEAT.0005664.

Torabi, M., Hashemi, S., Saybani, M.R., Shamshirband, S., Mosavi, A., 2019. A hybrid clustering and classification technique for forecasting short-term energy consumption. Environ. Prog. Sustain. Energy 38, 66–76. http://dx.doi.org/10.1002/ep.12934.

Walsh, J., 1981. Effect of pore pressure and confining pressure on fracture permeability. Int. J. Rock Mech. Min. Sci. Geomech. Abstr. 429–435. http://dx.doi.org/10.1016/0148-9062(81)90006-1, Elsevier.

Yin, Q., Yang, J., Li, Z., Huang, Y., Luo, M., Wang, B., Tyagi, M., Xu, G., Zhao, X., 2020. A field case study of managed pressure drilling in offshore ultra high-pressure high-temperature exploration well in the South China Sea. In: SPE Drilling and Completion. http://dx.doi.org/10.2118/191060-PA.

Zhang, G., Davoodi, S., Shamshirband, S., Ghorbani, H., Mosavi, A., Moslehpour, M., 2022. A robust approach to pore pressure prediction applying petrophysical log data aided by machine learning techniques. Energy Rep. 8, 2233–2247. http://dx.doi.org/10.1016/j.egyr.2022.01.012.