Novel hybrid machine learning optimizer algorithms to prediction of fracture density by petrophysical data

Meysam Rajabi¹ · Saeed Beheshtian² · Shadfar Davoodi³ · Hamzeh Ghorbani⁴ · Nima Mohamadian⁵ · Ahmed E. Radwan⁶ · Mehdi Ahmadi Alvar⁷

Received: 26 July 2021 / Accepted: 29 September 2021 / Published online: 9 October 2021
 © The Author(s) 2021

Abstract

One of the challenges in reservoir management is determining the fracture density (FVDC) in reservoir rock. Given the high cost of coring operations and image logs, the ability to predict FVDC from various petrophysical input variables using a supervised learning basis calibrated to the standard well is extremely useful. In this study, a novel machine learning approach is developed to predict FVDC from 12-input variable well-log based on feature selection. To predict the FVDC, combination of two networks of multiple extreme learning machines (MELM) and multi-layer perceptron (MLP) hybrid algorithm with a combination of genetic algorithm (GA) and particle swarm optimizer (PSO) has been used. We use a novel MELM-PSO/GA combination that has never been used before, and the best comparison result between MELM-PSO-related models with performance test data is RMSE = 0.0047 1/m; $R^2 = 0.9931$. According to the performance accuracy analysis, the models are MLP-PSO < MLP-GA < MELM-GA < MELM-PSO. This method can be used in other fields, but it must be recalibrated with at least one well. Furthermore, the developed method provides insights for the use of machine learning to reduce errors and avoid data overfitting in order to create the best possible prediction performance for FVDC prediction.

Keywords Fracture density · Multi-hidden layer extreme learning machine · Hybrid machine learning algorithms · Multi-layer perceptron

Abbreviations

AAPD Absolute average percentage deviation
 ANN Artificial neural network
 ANFIS Adaptive neuro fuzzy inference system
 APD Average percentage deviation
 CFD Cumulative distribution functions

CMIS Combination with the committee machine intelligent system
 CP Caliper
 CT Computed tomography
 $d_{\max}$ Maximum aggregate size
 DR Deep resistivity
 DT Sonic interval transit time

✉ Hamzeh Ghorbani
hamzehghorbani68@yahoo.com
 Meysam Rajabi
m.rajabi@birjandut.ac.ir
 Saeed Beheshtian
beheshtian.saeed@gmail.com
 Shadfar Davoodi
davoodis@hw.tpu.ru
 Nima Mohamadian
nima.0691@gmail.com
 Ahmed E. Radwan
radwanae@yahoo.com
 Mehdi Ahmadi Alvar
m-ahmadi-alvar@stu.scu.ac.ir

¹ Department of Mining Engineering, Birjand University of Technology, Birjand, Iran
² Department of Petroleum Engineering, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran
³ School of Earth Sciences and Engineering, Tomsk Polytechnic University, Lenin Avenue, Tomsk, Russia
⁴ Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
⁵ Young Researchers and Elite Club, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran
⁶ Faculty of Geography and Geology, Institute of Geological Sciences, Jagiellonian University, Gronostajowa 3a, 30-387 Kraków, Poland
⁷ Department of Computer Engineering, Faculty of Engineering, Shahid Chamran University, Ahvaz, Iran

E	Youngs modulus
FVDC	Fracture density
GA	Genetic algorithm
G_b	Best positions achieved globally
GR	Gamma ray
ILD	Deep induction log
LSSVM	Least-squares support-vector machine
MLP	Multi-layer perceptron
MSE	Mean square error
NFR	Naturally fractured reservoirs
NPHI	Neutron porosity
P_b	Best positions for individual particles
PD_i	Percentage deviation
PEF	Photoelectric absorption factor
P -wave	Compressional wave
$Q1$ – $Q15$	Characters of input variables
R	The dataset of data records
R^2	Coefficient of determination
RBF	Radial basis function
RE	Relative error
RHOB	Bulk density
RMSE	Root mean squared error
SD	Standard deviation
SVM	Support-vector machine
S -wave	Shear wave
$V_{\max}$	Maximum values to determine the quality of PSO
$V_{\min}$	Minimum values to determine the quality of PSO
w/c	Water–cement ratio
x	Data variable value range
X	The value of variable x in a specific data record
f^C	Compressive strength
ϑ_i^l	The value of attribute l for data record I
$\vartheta \min^l$	The minimum value of the attribute l among all the data records in the dataset
$\vartheta \max^l$	The maximum value of the attribute l among all the data records in the dataset

Introduction

Naturally fractured reservoirs (NFR) are one of the most economically important reservoirs globally, which has received much attention from petroleum engineers and geoscientists (Ja'fari et al. 2012; Nelson 1985; Ouenes 2000). Sarkheil et al (2009) argued that natural fractures in reservoir rock structure are one of the reasons for high gas production from gas fields. The importance of NFR results from the developing relationship between porosity and permeability, as well

as the relationship between fracture and high recovery factors in these reservoirs (Allan and Sun 2003; Nelson 2001; Wang and Sharma 2019; Warpinski et al. 2009).

Natural fractures are observed in many of the world's reservoirs, and this factor is significant in the oil and gas industry because it directly contributes to fluid flow. Fractures are a critical factor in reservoir development and management, and several methods for detecting them have been developed, including the use of petrophysical logs, mud loss history, core descriptions, well tests, and seismic data (Aguilera 2008; Dutta et al. 2007; Gale et al. 2007; Ja'fari et al. 2012; Suboyin et al. 2020; Thompson 2000; Tokhmechi et al. 2009). The reservoir's fracture is extremely complex, and it is greatly influenced by geological conditions such as (Sarkheil et al. 2013):

- Under lithostatic
- Fluid pressure
- Tectonic
- Thermal and other geological stresses such as uplifting
- Volcanoes
- Salt intrusion

Many direct and in-direct methods have been identified and used for FVDC detection from the past to the present (Ja'fari et al. 2012). The disadvantages of these methods are that they are very expensive and time-consuming, as well as having some practical limitations for example image log (based on FMI log) and coring (Ja'fari et al. 2012; Tokhmechi et al. 2009). In recent years, many researchers have used low-cost and time-consuming artificial intelligence methods to predict FVDC in oil and gas reservoir formations (Zazoun 2013). Boadu (1998) predicted FVDC using the artificial neural network (ANN) method using 3050 random data, including compressional wave (P -wave) and shear wave (S -wave). The results show that this algorithm fits the data properly and has high-performance accuracy (Boadu 1998). Ince (2004) successfully applied an ANN technique to 40 data records involving a two-parameter algorithm including CTODC (mm) (critical crack tip opening displacement) and K_{IC}^S (Mpa $\sqrt{m}$) (critical stress intensity factor based on two-parameter model) to predict concrete fracturing. Later, Sarkheil et al. (2009) used the developed nonlinear modeling and forecasting system to evaluate and predict the reservoir fracture network using 17 wells from the Tabnak hydrocarbon field in Iran associated with image loges (FMI) and core measurements. The correlations between observed and predicted failure densities for training, validation, and test data were 0.92, 0.86, and 0.88, respectively, according to their

findings. The aforementioned Sarkheil et al (2009) study provides fundamental information for scientific research in these areas as well as the assessment of hydrocarbon production. Jafari et al. (2012) used data from 15 wells in Iran's Marun field, including deep resistance, neutron porosity, and bulk density log data. Based on the application of adaptive neuro fuzzy inference system (ANFIS) model, they were able to predict FVDC (Ja'fari et al. 2012). One year later, Zazoun (2013) achieved acceptable results using the ANN method with six inputs: core depth, gamma ray (GR), sonic interval transit time (DT), caliper, neutron porosity (NPHI) and bulk density (RHOB), fracture density (FVDC). With an R^2 value of 0.812, their four-layered ANN model using a conjugate gradient descent training algorithm provided the best fracture-density prediction performance (Zazoun 2013). Nouri-Taleghani et al (2015) used a combination of three models, multi-layer perceptron (MLP), radial basis function (RBF), and least-squares support-vector machine (LSSVM), in conjunction with the committee machine intelligent system (CMIS), to create combined models MLP-CMIS,

RBF-CMIS, and LSSVM-CMIS in Marun oil field models to predict FVDC based on the six inputs variables including depth, caliper, neutron porosity (NPHI), photoelectric absorption factor (PEF), bulk density (RHOB), and deep induction log (ILD). The results of Nouri-Taleghani et al (2015) revealed that the LSSVM-CMIS hybrid model outperformed the other two hybrid models ($R^2=0.8950$ and $RMSE=0.102$) (Nouri-Taleghani et al. 2015). Three years later, Li et al (2018) presented a model for predicting FVDC using algorithm synthesis support-vector machine (SVM) with genetic algorithm (GA) based on acoustic waves input variables and data related to carbonate reservoir in the Ordos Basin (China), and they predicting a correlation coefficient of 0.7 (Li et al. 2018). In the same year, Bhattacharya and Mishra (2018) obtained absolute accuracy of 74.8% and 79.6%, using random forest (RF) and Bayesian network (BN) algorithms with GR, RHOB, and CP inputs from the Appalachian Basin formations (USA) (Bhattacharya and Mishra 2018). According to recent research, intelligent machine learning models for accurately estimating the FVDC have

Table 1 Implementation of some machine learning algorithms on prediction of FVDC

Author	Year	Region	No. of Data	Artificial intelligence method	Input Variables	Error
(Boadu 1998)	1998	–	3050	ANN	Compressional wave (P -wave), Shear wave (S -wave)	$RMSE=0.1198$
(Ince 2004)	2004	–	40	ANN	Compressive strength (f'_c), maximum aggregate size (d_{max}), water–cement ratio (w/c), Youngs modulus (E)	–
(Sarkheil et al. 2009)	2009	Iran	–	ANN	Azimuth, density, dip	$R^2=0.92, 0.86$ and 0.88 (train, validation and test)
(Ja'fari et al. 2012)	2012	Iran	–	ANFIS	Deep resistivity (RT), neutron porosity (NPHI) and bulk density (RHOB)	$R^2=0.92$, $RMSE=16.2$
(Zazoun 2013)	2013	Algeria	1349	ANN	Core depth, gamma ray (GR), sonic interval transit time (DT), caliper, neutron porosity (NPHI) and bulk density (RHOB)	$R^2=0.812$
(Nouri-Taleghani et al. 2015)	2015	Iran	395	MLP-CMIS, RBF-CMIS and LSSVM-CMIS	Depth, caliper, neutron porosity (NPHI), photoelectric absorption factor (PEF), bulk density (RHOB), and deep induction log (ILD)	Best result (LSSVM-CMIS) $RMSE=10.2$
(Li et al. 2018)	2018	China	–	SVM-GA	Acoustic waves	–
(Bhattacharya and Mishra 2018)	2018	USA	–	BN and RF	Gamma ray (GR), bulk density (RHOB) and caliper (CP)	$APD=74.8–79.6\%$

Fig. 1 Workflow of formation FVDC prediction based on hybrid machine-learning optimizer models

provided inexpensive, fast, and accurate solutions, as shown in Table 1.

Gap of knowledge and proposed novel model

Numerous techniques, such as increased fluid loss rates into fissures and fractures or complete circulation loss during drilling operations, have been introduced to identify fracture reservoirs based on accurate observations during drilling (Feng et al. 2016; Xu et al. 2019). In the mid-1980s, advances in dip meter technology and image logs enable the detection of features such as FVDC (Luthi 2001; Serra 1989; Toyobo et al. 2020). Many authors argue that they can predict fracture density (FVDC) using micro-resistivity images (FMI), and the output of this log is image loge (Zellou and Ouenes 2003). Furthermore, image logs from advanced devices such as measurement-while-drilling (MWD) during drilling operations have been able to identify and determine fractures in reservoir rocks (Khorzoughi et al. 2018; Murphy 1993; Xu et al. 2016). Prior to the image logs technology, the fracture detection operation was performed in the laboratory using coring and X-rays, or computed tomography (CT) scans on the core, which both operations (image log and coring) required a significant cost and time (AlAwad and Fattah 2017; Andersen et al. 2013; Kuramoto et al. 2008; Romano et al. 2019; Zazoun 2013; Zerrouki et al. 2014). As a result, it will be highly beneficial to employ new techniques such as hybrid machine learning in order to obtain more accurate results while saving money and time.

This study attempted to provide a novel model (MELM with PSO/GA optimizer) that minimizes RMSE based on a database of more than 3019 data points, using 12 input variables based on feature selection, which includes: bulk formation density (RHOB); density correction (DRHO); thorium/uranium ratio (TURT); caliper (CALI); shallow resistivity (HMRS); thorium (THOR); the photoelectric index (PEF); neutron porosity (NPHI); potassium (POTA); deep resistivity (HDRS); uranium/potassium ratio (UKRT); and gamma ray (GR). Furthermore, several control measures are being used to increase accuracy, reduce errors, and avoid data overfitting in order to create the best possible prediction performance for FVDC prediction, which has never been done before in the literature.

Methodology

Workflow diagram

Four hybrid machine learning algorithms with two efficient optimizers are used in this paper to predict FVDC for three wells in the large Marun oil and gas field in southwestern Iran. Figure 1 shows the applied workflow diagram for predicting the FVDC. Based on this diagram, the data collection process was first applied to the studied field data from the Asmari carbonate reservoir. The data are then sorted to describe the data variable, and maximum and minimum values for each variable are determined. After normalizing the input data (Eq. 1), the feature selection

process is performed to determine a suitable combination of inputs. After determining the best combination of inputs, the input data related to the two wells are divided into three sections: test, train, and validation.

$$\mathfrak{z}_i^l = \left(\frac{\mathfrak{z}_i^l - \mathfrak{z} \min^l}{\mathfrak{z} \max^l - \mathfrak{z} \min^l} \right) * 2 - 1 \quad (1)$$

where $\mathfrak{z} \min^l$ and $\mathfrak{z} \max^l$ are the minimum and maximum values of the attribute l in an arrangement; $\mathfrak{z}_i^l$ is the value of attribute l for data record I .

One way to compare algorithms is to have a proper comparison (statistical parameters) (Hazbeh et al. 2021a). At this stage, a comparison is made based on statistical parameters, and after identifying the best algorithm, it is tested for development on another well.

Feature selection

One method for shortening the process and better estimating each model is to use the best features rather than all features, which increases the program's speed, efficiency, and accuracy (Farsi et al. 2021). This selection includes all of the features available for determining the best performance estimate (Jain and Zongker 1997). When the number of input variables is large, the probabilistic estimation becomes too repetitive and difficult to repeat consecutive performances. For example, if 15 attributes are available, there are as many as 2^N possible combinations (32,768) (Chandrashekar and Sahin 2014). There are numerous feature selection methods (for example, filtering, wrapping, and embedded methods), the most basic of which is the filtering method, and not every evaluated subset is optimal (John et al. 1994). Wrapping methods are more accurate and effective than other methods because they use evolutionary algorithms (e.g., genetic algorithm (GA)) to identify ineffective variables and eliminate potential properties. Variables are generated using potential solutions; each data set is then normalized, and the results are obtained using the cost function, which in this case is the root mean squared error. GA transfers high-performance solutions (minimum RMSE) to the next optimizer iteration and applies the main reader, which includes crossover, combination, and mutation, to the new iterations (Wahab et al. 2015).

Machine learning algorithms

Many researchers have conducted extensive research to determine the key factors in many areas of the oil and gas industry, including formation damage (Mohammadian and Ghorbani 2015), CO₂ capture (Hassanpouryouzband et al. 2019; Hassanpouryouzband et al. 2018a, b),

reservoir performance (Ghorbani et al. 2017a), wellbore stability (Darvishpour et al. 2019), production (Ghorbani and Moghadasi 2014; Ghorbani et al. 2017b, 2014), rheology and filtration (Mohamadian et al. 2018), wellbore blowout (Abdali et al. 2021) and drilling fluid optimization (Mohamadian et al. 2019). Many studies in machine learning and new algorithms can aid in the solving of technical problems in many areas of the oil and gas industry (Choubineh et al. 2017; Ghorbani et al. 2020). In this study, four hybrid algorithms combine two networks of multi-extreme learning machines (MELM) and MLP using particle swarm optimization (PSO) and GA, including MELM-PSO, MELM-GA, MLP-PSO, and MLP-GA.

Multi-layer perceptron

Artificial neural networks (ANNs) are one method for facilitating accurate prediction of dependent variables and complex methods and equations (Ali 1994). Because of the variety of complexities for each dependent variable, there are many different types of neural networks. The selection of attributes (i.e., the input variables to be considered), network architecture (number of layers and nodes), and the transfer of functions between layers, as well as the selection of training algorithm, are among the factors that increase the correct choice of these factors and thus increase the performance accuracy in artificial networks (Ghorbani et al. 2018; Mohamadian et al. 2021). The multi-layer perceptron (Fig. 2) is one of the most practical and adaptable neural networks for large and complex datasets (Rashidi et al. 2020). Therefore, MLP is very useful and suitable for predicting FVDC. Combining networks with different evolutionary algorithms is one way to improve performance and results in network algorithms. To improve the results, the MLP methodology employs two evolutionary genetic algorithms (GA) and one particle

Fig. 2 A schematic of MLP by showing input, hidden and output layer and nodes

swarm algorithm (PSO). One of the algorithms in the train that helps the MLP algorithm in fast convergence to the optimizer is the Levenberg–Marquardt (LM) algorithm.

Multi-layer extreme learning machine

ELM, introduced by Liang et al (2006), is one of the fastest and easiest computing networks, avoiding time-consuming repetitions during ELM network training with single hidden layers (Liang et al. 2006). The input weight/bias in this network is selected at random from a uniform distribution and is not usually adjusted through network tuning (Fig. 3) (Cheng and Xiong 2017; Huang et al. 2011; Liang et al. 2006). The output weights of the ELM are converted to a linear output layer by inverting the hidden layer, and the optimal value of the output weight is identified using root mean-squared error regression (Huang et al. 2011). This eliminates the need for duplicate backpropagation simulation, which is an essential part of MLP networks (Yeom and Kwak 2017). The use of latent multi-layer ELM (MELM) is common in complex nonlinear systems that require classification. This new network (MELM) can achieve higher accuracy and generalizability than concealed single-layer models. The structure of MELM can be shown in (Rashidi et al. 2021).

Genetic algorithm

The genetic algorithm (GA) is one of the evolutionary algorithms that simulates feature selection and a suitable combination for hybrids with MLP and MELM and solves problems repeatedly (Simon 2013). One of the high-performance solutions is identified in each iteration and preferably used to help make changes to new solutions for the next GA iteration, while weaker performance solutions based on their poor fitness comparison are gradually eliminated (Mirjalili 2019).

Fig. 3 A schematic of the simple network structure of ELM

Particle swarm optimization algorithm

The PSO algorithm is a common evolutionary algorithm based on the crowding of groups. Each particle has the potential to solve the problem, and the particle swarm represents one possible solution to the problem (Eberhart and Kennedy 1995). The PSO algorithm uses iteration to find the best possible random solution and to set the minimum ($V_{\min}$) and maximum ($V_{\max}$) values to determine the quality. Best positions for individual particles (P_b) and best positions achieved globally (G_b) by the entire particle swarm are recorded for each iteration of the algorithm (Atashnezhad et al. 2014). All information is transferred to the next iteration, and the position of each particle is adjusted with its velocity and the lowest G_b and regarding the lowest P_b of the particles of each particle is done, the velocity V_t sets them to V_{t+1} , where t refers to the number of iterations. For each iteration, the particles in the swarm reach the lowest RMSE target performance values that have the greatest impact on the next generation of particles (Rashidi et al. 2021). References to related velocity changes and flowcharts can be found in the articles of Singh et al. (2020) and Cai et al. (2020).

Hybrid MELM-PSO/GA model

One of the initial requirements of MELM neural network is to determine the number of hidden layers (l) and the number of neural cells (n) related to each outset, which is usually determined by trial and error (Tang et al. 2014). One of the benefits of using an optimizer algorithm is that it increases the speed and scope of the search for practical values. Finding these optimal parameters in network structure formation is very sensitive and important because selecting these optimal parameters can increase the number of hidden layers, hidden neurons, and nodes in these layers, resulting in structural complexity and inefficiency of the model, as well as increasing run time and preventing network structure formation (Zhang et al. 2016). In this paper, we use one of the most recent MELM fabrication methods (Fig. 4) to help identify the desired number of hidden layers and nodes in these layers (as opposed to trial and error); and after establishment, to help identify the optimal values of weights to be applied to each neuron in each latent layer, as well as the biases applied to each latent layer (replacement of random assignment of these values) (Farsi et al. 2021; Su et al. 2019; Zheng et al. 2019). In this article, a hybrid algorithm that combines PSO and GA optimization algorithms with a MELM network is used, which has not previously been used in this field. The combination of PSO with MELM forms a combination of PSO-MELM-PSO, and the combination of

Fig. 4 Flow chart for hybrid MELM-PSO/GA model to prediction of fracture detection

Table 2 Report of RMSE value for the numbers of neurons in the layers to MELM network for the prediction of FVDC

Number of hidden layers	Number of neurons in the layers				
	5	10	15	20	25
3	0.13229	0.12933	0.12819	0.12875	0.12947
5	0.13005	0.12769	0.12158	0.12691	0.12751
7	0.13021	0.12932	0.12236	0.12913	0.13023
9	0.13042	0.13009	0.12977	0.12986	0.13074

GA with this network makes GA-MELM-GA. In this study, the number of hidden layers of MELM is allowed to vary from 3 to 9, and the number of nerve cells in each hidden layer is allowed to vary from 5 to 25 in order to determine and predict FVDC (Table 2). MELM-PSO has 3 to 7 hidden layers and 10 to 20 neurons for each hidden layer with higher accuracy (lower RMSE), as shown in Table 2, and

control parameters used for predicting fracture detection for the algorithm specify PSO and GA in Tables 3 and 4.

Hybrid MLP-PSO/GA model

A schematic of how to implement the MLP-PSO and MLP-GA models (Hazbeh et al. 2021b) is shown in Fig. 5. To optimize the weight and layer nodes (or neurons) in these hybrid neural networks, the MLP network is hybridized with GA and PSO algorithms (Rashidi et al. 2021; Sabah et al. 2019), as shown in Tables 5 and 6 of the control values, the PSO and GA parameters are given.

Statistical errors for FVDC prediction

To predict FVDC using HML and statistical errors, a comparison of artificial intelligence methods such as mean square error (MSE), percentage deviation (PD_i), relative error (RE), average percentage deviation (APD), absolute average percentage deviation (AAPD), coefficient of

Table 3 Parameter's control for the PSO algorithms to PSO-MELM-PSO for the prediction of FVDC

MELM		PSO	
Parameters' control	Values	Parameters' control	Values
Number of inputs	7	Maximum number of iterations	100
Number of neurons in the layers	20	Swarm size	60
Number of hidden layers	5	Cognitive constant	2.05
Performance	RMSE	Social constant	2.05
		Inertia weight (Damping ratio)	0.99
		Var Min	−10
		Var Max	10
		Intertie coefficient (w)	1
		Max velocity	2
		Min velocity	−2

Table 4 Parameter's control for the GA algorithms to GA-MELM-GA for the prediction of FVDC

MELM		GA	
Parameters' control	Values	Parameters' control	Values
Number of inputs	15	Population	60
Number of neurons in the layers	20	Selection method	Roulette wheel
Number of hidden layers	5	Crossover	Uniform ($p = 1$)
Performance	RMSE	Mutation	Uniform ($p = 0.04$)
		Mutation rate	0.11
		Selection pressure (Roulette wheel)	2

determination (R^2), root mean square error (RMSE; the objective function of the HML models), and standard deviation (SD). The equations used in this work are given as Eqs. (2, 3, 4, 5, 6, 7, 8) in the following section:

Mean Square Error (MSE):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Z_{\text{Measured } i} - Z_{\text{Predicted } i})^2 \quad (2)$$

Percentage deviation (PD_i) or relative error (RE):

$$\text{PD}_i = \frac{H_{(\text{Measured})} - H_{(\text{Predicted})}}{H_{(\text{Measured})}} \times 100 \quad (3)$$

Average percentage deviation (APD):

$$\text{APD} = \frac{\sum_{i=1}^n \text{PD}_i}{n} \quad (4)$$

Absolute average percentage deviation (AAPD):

$$\text{AAPD} = \frac{\sum_{i=1}^n |\text{PD}_i|}{n} \quad (5)$$

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^N (H_{\text{Predicted } i} - H_{\text{Measured } i})^2}{\sum_{i=1}^N \left(H_{\text{Predicted } i} - \frac{\sum_{i=1}^N H_{\text{Measured } i}}{n} \right)^2} \quad (6)$$

Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (7)$$

Standard Deviation (SD):

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^n (D_i - D_{\text{mean}})^2}{n - 1}} \quad (8)$$

$$D_{\text{mean}} = \frac{1}{n} \sum_{i=1}^n (H_{\text{Measured } i} - H_{\text{Predicted } i})$$

Fig. 5 Flow chart for hybrid MLP-PSO/GA model to prediction of fracture detection

Data collection, description, and distribution

Data collection

The Aghajari oil field, which includes the Asmari and Bangestan formations, is one of the oldest discovered fields and the largest in the Zagros basin. To investigate and predict

FVDC, information from three wells AJ#A, AJ#B, and AJ#C in the huge Aghajari oil field on Iran's southwest coast was used (Fig. 6). There are 1026 datasets related to information collected from well AJ#A, 1057 datasets related to information collected from well AJ#B, and 936 datasets related to information collected from well AJ#C.

Data description

In this work, well AJ#C was used to develop and generalize this model in other wells in order to train the software and based on the data diversity of wells AJ#A and AJ#B. The total employed data include 1026 data from well AJ#A with a distance of 3824–4131 m, 1057 data from well AJ#B with a distance of 3801–4118 m, and 936 data from well AJ#C with a distance of 3905–4185 m. The data record distance for the data sets for each well is 0.3 m. Common well-log and wireline tools record petrophysical data (Ghasemi and Bayuk 2020). Some of the petrophysical data used in this study to predict FVDC include: bit size (BS); caliper (CALI); density correction (DRHO); neutron porosity (NPHI); the photoelectric index (PEF); potassium (POTA); uranium (URAN); thorium (THOR); thorium/potassium ratio (TKRT); thorium/uranium ratio (TURT); uranium/potassium ratio (UKRT); bulk formation density (RHOB); gamma ray (GR); deep resistivity (HRSR); and shallow resistivity (HMRs). Tables 7 and 8 show statistical characterization for whole and individual data sets (AJ#A = 1026 data records; AJ#B = 1057 data records; AJ#C = 936 data records), respectively.

Table 5 Parameter's control for the PSO algorithms to PSO-MELM-PSO for the prediction of FVDC

MLP		PSO	
Parameters' control	Values	Parameters control	Values
Number of inputs	15	Maximum number of iterations	100
Number of hidden layers	2	Swarm size	60
Number of input layer neurons	7	Cognitive constant	2.05
Number of hidden layer1 neurons	10	Social constant	2.05
Number of hidden layer2 neurons	5	Inertia weight (Damping ratio)	0.99
Number of output layer neurons	1	Var Min	− 10
Performance	RMSE	Var Max	10
		Intertie coefficient (w)	1
		Max velocity	2
		Min velocity	− 2

Table 6 Parameter's control for the GA algorithms to GA-MELM-GA for the prediction of FVDC

MLP		GA	
Parameters' control	Values	Parameters' control	Values
Number of inputs	15	Population	60
Number of hidden layers	2	Selection method	Roulette wheel
Number of input layer neurons	15	Crossover	Uniform ($p = 1$)
Number of hidden layer1 neurons	10	Mutation	Uniform ($p = 0.04$)
Number of hidden layer2 neurons	5	Mutation rate	0.11
Number of output layer neurons	1	Selection pressure (Roulette wheel)	2
Performance	RMSE		

Data distribution

The cumulative distribution functions (CDF) (Hazbeh et al. 2021a) are one of the criteria used to describe the data (Eq. 9).

$$F_X(x) = P(X \leq x), \text{ for all } x \in R \quad (9)$$

where $F_X(x)$ is cumulative distribution functions; x is data variable value range; X is the value of variable x in a specific data record; and R is the dataset of data records.

Table 9 provides useful visually statistical information of variables for Figs. 7 and 8 (CFD for input variables) hairs with normal distributions defined by variable means and standard deviations (thicker red line).

Based on the CFDs shown in Figs. 7 and 8, it can be seen that the information about the seven parameters POTA, TKRT, TURT, UKRT, HDRS, HMRS, and FVDC is not normally distributed, but the other parameters except BS that all data are the same and are normally distributed. After examining Fig. 8, it can be seen that it has the largest deviation from the normal HDRS distribution.

Fig. 6 Schematic of Aghajari oil field's location in southwest Iran

Result and discussion

Feature selection for FVDC prediction

One of the most important factors for improving performance and high accuracy in optimizing hybrid models is feature selection after constructing the MLP-GA model and finding the reducing of the number of independent input variables in the models using multiple evaluations and training, as well as finding the MLP structure. Based on the identified two layers with 6 and 5 nodes in their first and second hidden layers, the most effective feature selection model in minimizing RMSE can predict FVDC values. Using the MLP architecture, the approach randomly selected 30% of the total data records available for the experimental subset before assigning the remaining 70% to the training subset. Given that this method was unable to prevent the increase in the feature selection process, some features have a disproportionate effect on the predictions. The feature selection process is as follows: First, the features are selected into one, double, triple, etc., categories up to fifteen. In the next step, for example, in categories one, enter each feature as input to the MLP-GA algorithm and after outputting the RMSE value is obtained for them. Finally, for one category, the best RMSE is reported in Table 5, which here is HDRS (Q14). This comparison shows us that the most effective input feature among the categories is HDRS, and the same is true for the other categories. Table 10 implements a method for determining the combination of 15 variables in feature selection. In order to determine the combination of 15 variables in the feature selection in Table 10 according to Q1–Q15.

After checking the feature selection and combining several inputs, it calculates the RMSE values according to the selection of different inputs. Following a review, it was determined that the 12 input variables that outperformed on

Table 7 Statistical characterization to prediction of FVDC prediction of 3019 data variables to the well-log dataset for all Aghajari oil field wells

Units	Vari- ables	Bit Size	Caliper	Density correc- tion	Neutron porosity	The photo- electric index	Potas- sium	Ura- nium	Tho- rium	Tho- rium/ potas- sium ratio	Tho- rium/ uranium ratio	Ura- nium/ potas- sium ratio	Bulk density	Gamma ray	Deep resis- tivity	Shallow resistivity	Fracture density
Symbol	BS	CALI	DRHO	NPHI	PEF	POTA	URAN	THOR	TKRT	TURT	UKRT	RHOB	GR	HDRS	OHMM	OHMM	FVDC
Units	Inch	Inch	g/cm3	PU	Barn/ cm3	PPM	PPM	PPM	PPM	–	–	–	g/cm3	GAPI	OHMM	OHMM	1/m
3019 dataset records from AI#A, AI#B and AI#C	Mean	5.875	5.951	0.024	0.062	4.783	2.989	1.929	1.124	0.943	2.468	1.725	2.618	21.673	3676.36	3649.67	0.667
	Std. Devia- tion	0.000	0.300	0.041	0.045	0.214	3.760	1.135	0.799	1.223	9.544	2.423	0.083	11.953	7494.94	7408.55	1.484
	Variance	0.000	0.090	0.002	0.002	0.046	14.135	1.288	0.638	1.495	91.065	5.870	0.007	142.819	5.62E+07	5.49E+07	2.202
	Mini- mum	5.875	5.525	0.000	0.005	3.405	0.086	0.034	0.089	0.032	0.028	0.014	2.345	2.219	0.001	0.728	0.000
	Maxi- mum	5.875	8.556	0.634	0.212	5.961	31.011	8.100	5.316	10.194	99.966	17.570	2.962	79.010	2.25E+04	2.51E+04	10.431
	Skew- ness	NA	4.271	4.360	0.987	–1.211	2.830	0.719	2.493	2.781	6.830	2.683	–0.734	0.961	1.675	1.693	2.686
	Kurtosis	NA	24.926	30.762	0.026	4.909	9.158	1.460	7.136	8.999	48.940	7.494	0.073	1.090	0.864	0.950	8.373

Table 8 Statistical characterization to prediction of FVDC prediction of the data variables to the well-log dataset for three Aghajari oil field wells: AJ#A, AJ#B and AJ#C

Units	Variables	Bit Size	Caliper	Density correction	Neutron porosity	The photo-electric index	Potassium	Uranium	Thorium	Thorium/potassium ratio	Thorium/uranium ratio	Uranium/potassium ratio	Bulk density	Gamma ray	Deep resistivity	Shallow resistivity	Fracture density
	Symbol	BS	CALI	DRHO	NPHI	PEF	POTA	URAN	THOR	TKRT	TURT	UKRT	RHOB	GR	HDRS	OHMM	FVDC
	Units	Inch	Inch	g/cm3	PU	Barn/cm3	PPM	PPM	PPM	–	–	–	g/cm3	GAPI	OHMM	OHMM	1/m
1026 dataset records from AJ#A	Mean	5.875	5.925	0.018	0.052	4.807	1.375	1.332	0.855	1.284	2.668	1.896	2.619	14.280	3986.41	4411.69	0.820
	Std. Deviation	0.000	0.300	0.034	0.034	0.204	0.989	0.893	0.311	1.476	5.457	2.441	0.069	7.460	7756.33	7984.65	1.718
	Variance	0.000	0.090	0.001	0.001	0.042	0.977	0.797	0.097	2.175	29.750	5.952	0.005	55.603	6.01E+07	6.37E+07	2.949
	Minimum	5.875	5.678	0.000	0.005	3.705	0.086	0.034	0.139	0.060	0.050	0.014	2.444	2.219	0.001	1.482	0.000
	Maximum	5.875	8.347	0.634	0.165	5.961	5.512	3.528	2.173	10.194	31.108	12.734	2.962	37.422	2.25E+04	2.21E+04	10.431
	Skewness	N/A	5.652	10.290	0.518	–1.528	0.916	0.098	0.683	2.310	3.233	2.140	–0.012	0.160	1.547	1.392	2.653
	Kurtosis	N/A	38.177	151.232	–0.745	5.549	0.685	–1.123	0.991	5.526	9.964	4.029	–0.277	–0.937	0.441	0.003	8.276
1057 dataset records from AJ#B	Mean	5.875	5.821	0.017	0.054	4.767	4.344	2.481	1.326	0.763	0.687	1.896	2.653	28.511	3637.53	3306.49	0.614
	Std. Deviation	0.000	0.079	0.024	0.039	0.206	4.798	1.173	0.917	1.016	0.753	2.711	0.060	13.279	7404.55	7019.93	1.417
	Variance	0.000	0.006	0.001	0.002	0.042	22.995	1.374	0.840	1.032	0.567	7.340	0.004	176.171	5.48E+07	4.92E+07	2.005
	Minimum	5.875	5.639	0.000	0.005	4.008	0.153	0.119	0.103	0.032	0.028	0.058	2.416	5.286	3.445	4.790	0.000
	Maximum	5.875	6.507	0.220	0.212	5.149	23.074	8.100	5.316	9.526	11.918	17.570	2.813	79.010	2.20E+04	2.03E+04	8.492
	Skewness	N/A	2.081	4.795	1.513	–1.094	1.695	0.940	1.554	3.205	6.226	2.571	–1.073	0.677	1.702	1.861	2.566
	Kurtosis	N/A	11.567	25.004	2.040	0.979	2.239	1.433	2.153	13.882	69.504	6.715	1.738	–0.120	0.974	1.591	6.675

Table 8 (continued)

Units	Vari- ables	Bit Size	Caliper	Density correc- tion	Neutron porosity	The photo- electric index	Potas- sium	Ura- nium	Tho- rium	Tho- rium/ potas- sium ratio	Tho- rium/ uranium ratio	Ura- nium/ potas- sium ratio	Bulk density	Gamma ray	Deep resis- tivity	Shallow resistivity	Fracture density
	Symbol	BS	CALI	DRHO	NPHI	PEF	POTA	URAN	THOR	TKRT	TURT	UKRT	RHOB	GR	HDRS	HMRS	FVDC
	Units	Inch	Inch	g/cm3	PU	Barn/ cm3	PPM	PPM	PPM	–	–	–	g/cm3	GAPI	OHMM	OHMM	1/m
936 dataset records from AI#C	Mean	5.875	6.127	0.039	0.081	4.773	3.226	1.959	1.190	0.771	4.259	1.345	2.578	22.053	3380.36	3201.93	0.560
	Std. Devia- tion	0.000	0.367	0.055	0.055	0.230	3.685	0.998	0.936	1.039	15.935	1.979	0.101	9.380	7297.26	7116.47	1.252
	Variance	0.000	0.135	0.003	0.003	0.053	13.568	0.995	0.874	1.078	253.638	3.911	0.010	87.898	5.32E+07	5.06E+07	1.565
	Mini- mum	5.875	5.525	0.000	0.011	3.405	0.122	0.046	0.089	0.045	0.035	0.031	2.345	5.089	0.703	0.728	0.000
	Maxi- mum	5.875	8.556	0.273	0.202	5.665	31.011	6.447	4.996	7.997	99.966	15.492	2.792	63.576	2.21E+04	2.51E+04	6.340
	Skew- ness	N/A	2.877	1.926	0.376	–1.057	3.303	0.557	2.541	2.858	4.259	3.658	–0.292	0.828	1.804	1.905	2.430
	Kurtosis	N/A	11.376	2.881	–1.398	7.407	13.863	1.574	6.513	8.947	16.463	15.286	–1.084	1.117	1.305	1.696	5.613

Table 9 Description of input variables for FVDC prediction is to use cumulative distribution functions (CDFs) for three Aghajari oil field wells

Parameters	Symbol	Unit	Boundary	Percent (%)
Bit size	BS	Inch	BS = 5.875	= 100
Caliper	CALI	Inch	< 5.82	~ 33.5
			5.82 < CALI < 6.44	~ 62.5
			CALI > 6.44	~ 4
Density correction	DHRO	g/cm ³	DHRO < 0.0104	~ 36.5
			0.0104 < DHRO < 0.0779	~ 55.1
			DHRO > 0.0779	~ 8.4
Neutron porosity	NPHI	PU	NPHI < 0.0213	~ 17.93
			0.0213 < NPHI < 0.1033	~ 62.5
			NPHI > 0.1033	~ 19.76
The photoelectric index	PEF	Barn/cm ³	PEF < 4.48	~ 7.62
			4.48 < PEF < 4.88	~ 60.7
			PEF > 4.88	~ 31.8
Potassium	POTA	PPM	POTA < 1.148	~ 30.04
			1.148 < POTA < 7.791	~ 60.88
			POTA > 7.791	~ 9.08
Uranium	URAN	PPM	URAN < 1.735	~ 47.3
			1.735 < URAN < 3.733	~ 46.07
			URAN > 3.733	~ 6.63
Thorium	THOR	PPM	THOR < 0.73	~ 30.2
			0.73 < THOR < 2.17	~ 61
			THOR > 6.44	~ 8.80
Thorium/potassium ratio	TKRT	–	TKRT < 0.3240	~ 30.34
			0.3240 < TKRT < 2.48	~ 59.31
			TKRT > 2048	~ 10.35
Thorium/uranium ratio	TURT	–	TURT < 0.4784	~ 40.77
			0.4784 < TURT < 18.89	~ 56.03
			TURT > 2048	~ 3.2
Uranium/potassium ratio	UKRT	–	UKRT < 0.5304	~ 30.4
			0.5304 < UKRT < 5.267	~ 60.1
			UKRT > 5.267	~ 9.5
Bulk density	RHOB	g/cm ³	RHOB < 2.55	~ 21.46
			2.55 < RHOB < 2.69	~ 61.18
			RHOB > 2.69	~ 17.36
Gamma ray	GR	GAPI	GR < 13.36	~ 22.12
			13.36 < GR < 35.5	~ 65.65
			GR > 35.5	~ 12.2
Deep resistivity	HDRS	OHMM	HDRS < 17.0	~ 33.35
			17.0 < HDRS < 10,200.7	~ 48.86
			HDRS > 10,200	~ 17.79
Shallow resistivity	HMRS	OHMM	HMRS < 16.1	~ 29.50
			0.4784 < HMRS < 10,954.9	~ 53.2
			HMRS > 10,954.9	~ 3.2
Fracture density	FVDC	1/m	FVDC < 0.3124	~ 30.42
			0.3124 < FVDC < 2.65	~ 56.72
			FVDC > 2.65	~ 12.86

Fig. 7 The input variables and output values for the fracture prediction based on the CDF includes: BS (inch), CALI (inch), DRHO (g/cm³), NPHI (PU), PEF (Barn/cm³), POTA (PPM), URAN (PPM), THOR (PPM). In the cumulative distribution function (CDF), the

input variables and output (thinner blue line) compared to cumulative distribution functions for normal distributions defined by the variable means and standard deviations (thicker red line)

feature selection included RHOB, DRHO, TURT, CALI, HMRS, THOR, PEF, NPHI, POTA, HDRS, UKRT, and URAN (based on Table 11 and Fig. 9).

Results for FVDC prediction

The RMSE value is one of the best and most useful statistical errors for predicting the performance of any algorithm. This algorithm was chosen because it minimizes the target performance, which four models minimize, and better performance accuracy equals less RMSE. According to 2083 data, the statistical information measured to predict FVDC in Tables 12 13 14 includes four hybrid machine learning models developed, which include MLP-GA, MLP-PSO, MELM-GA, and

MELM-PSO models. The 2083 data points are associated with two wells, AJ#A and AJ#B, and are compared using a 12-input variable feature selection method. From 2083 data related to two wells A and B, 1463 data records; ~70% related to training subset, 310 Data Records; ~15% related to testing subset and 310 data records; ~15% is related to validation subset.

Using the results shown in Tables 12 13 14, it is clear that MELM-PSO has higher performance accuracy than other HML algorithms for predicting FVDC. The new and hybrid MELM-PSO algorithm has the best answer in this field for training, test, and validation subsets, as shown in Tables 12, 13, 14, with $RMSE = 0.0053$ 1/m; $R^2 = 0.9903$ (for training

Fig. 8 The input variables and output values for the fracture prediction based on the CDF include: TKRT (-), TURT (-), UKRT (-), RHOB (g/cm³), GR (GAPI), HDRS (OHMM) and HMRS (OHMM), FVDC (1/m). In the cumulative distribution function (CDF), the

input variables and output (thinner blue line) compared to cumulative distribution functions for normal distributions defined by the variable means and standard deviations (thicker red line)

subset); RMSE = 0.00401/m; $R^2 = 0.9959$ (for testing subset) and RMSE = 0.0047 1/m; $R^2 = 0.9931$ (for validation subset). Figure 10 compares the predicted FVDC to the measured FVDC for each data record in the training, testing, and validation subset for the four MLP-GA, MLP-PSO, MELM-GA, and MELM-PSO models. Evaluations have shown that combining optimal PSO and GA algorithms improves MELM performance more than combining these algorithms with an MLP network. The model compares and arranges MELM-PSO > MELM-GA > MLP-GA > MLP-PSO based on the correlation coefficients in Tables 12, 13, 14, and Fig. 10.

Figure 11 shows FVDC prediction error histograms displayed with normal distributions (red line) for HML algorithms based on 2083 subset data records from the Aghajari oil fields (AJ#A and AJ#B). As shown in Fig. 11, the best error for these models is MELM-PSO, and the lowest FVDC prediction error range is associated with is MELM-PSO model.

Figure 12 shows the RMSE diagram versus the repetition number for the four hybrid machine learning algorithms model MLP-GA, MLP-PSO, MELM-GA, and MELM-PSO for 100 replications. As shown in Fig. 12, MLP-GA,

Table 10 Description of input variables is to use cumulative distribution functions (CDFs) for three Aghajari oil field wells

Parameter	Unit	Character
BS	Inch	Q1
DRHO	g/cm ³	Q2
POTA	PPM	Q3
NPHI	PU	Q4
CALI	Inch	Q5
PEF	Barn/cm ³	Q6
URAN	PPM	Q7
THOR	PPM	Q8
TKRT	–	Q9
TURT	–	Q10
UKRT	–	Q11
RHOB	g/cm ³	Q12
GR	GAPI	Q13
HDRS	OHMM	Q14
HMRS	OHMM	Q15

MLP-PSO, and MELM-GA algorithms converge in iterations 3, 98, and 11, respectively, but for the MELM-PSO algorithm, it also converges at the beginning. As shown in the zoomed section of Fig. 12, it is well shown from the 50 iterations that the performance accuracy of the models is MLP-PSO < MLP-GA < MELM-GA < MELM-PSO, respectively.

Development and generalization of MELM-PSO model

According to the information related to the two wells AJ#A (1026 datasets) and AJ#B (1057 datasets) and the results related to all four models used in this research to predict FVDC (MLP-GA, MLP-PSO, MELM-GA, and MELM-PSO) and presenting the results in Tables 12, 13, 14, the MELM-PSO model has better performance accuracy than other models (based on 12-input variable feature selection). This model can be concluded by using well AJ#C (936 datasets) that the MELM-PSO model has been used for other wells in these fields and has a favorable result (Table 15) (RMSE = 0.0041 1/m; $R^2 = 0.9964$).

Figure 13 shows the predicted FVDC versus the measured value for the AJ#C well when developing the MELM-PSO model. Given the algorithm's performance accuracy, it is logical to say that it can be used in the Aghajari field. This method can be used in other fields, but it must be recalibrated with at least one well from another field before it can be used in another field.

Conclusions

Large data sets of 3019 data points from three wells in the Aghajari field (AJ#A, AJ#B, and AJ#C) in southwestern Iran are used to predict FVDC using four recombinant hybrid algorithms: MLP-GA, MLP-PSO, MELM-GA, and MELM-PSO. 2083 datasets from wells AJ#A and AJ#B were used to train the supervised algorithms, 1463 data

Table 11 Feature selection for FVDC prediction validation MLP-GA model to find best optimize

Number of input variables	Input variables	RMSE (1/m)
1	Q14	1.56263
2	Q6, Q14	1.35046
3	Q14, Q4, Q6	1.14438
4	Q4, Q6, Q14, Q11	0.96192
5	Q10, Q11, Q4, Q6, Q14	0.8134
6	Q11, Q3, Q6, Q10, Q14, Q4	0.75357
7	Q3, Q2, Q6, Q14, Q4, Q11, Q10	0.72325
8	Q5, Q6, Q14, Q3, Q2, Q10, Q4, Q11	0.6823
9	Q6, Q5, Q2, Q11, Q14, Q3, Q4, Q10, Q8	0.63413
10	Q14, Q2, Q11, Q6, Q8, Q3, Q4, Q10, Q5, Q15	0.59428
11	Q3, Q13, Q2, Q11, Q4, Q6, Q8, Q5, Q14, Q15, Q10	0.58186
12	Q13, Q2, Q10, Q5, Q15, Q8, Q6, Q4, Q3, Q14, Q11, Q7 (Best)	0.56783
13	Q4, Q3, Q10, Q7, Q11, Q14, Q6, Q9, Q2, Q8, Q15, Q5, Q13	0.59498
14	Q13, Q5, Q3, Q11, Q8, Q10, Q14, Q9, Q6, Q12, Q7, Q15, Q4, Q2	0.60178
15	Q1, Q2, Q5, Q6, Q9, Q8, Q4, Q7, Q11, Q3, Q12, Q14, Q15, Q10, Q13	0.62526

Fig. 9 Feature selection for FVDC prediction to optimize the best input variable

sets were used for training, 310 for testing, and 310 for validation. We used feature selection to avoid unnecessary inputs that reduce performance accuracy and to find the best inputs about each other to achieve the best result in terms of performance accuracy (MLP -GA). Based on feature selection, twelve input variable well-logs are considered: bit size (BS); caliper (CALI); density correction (DRHO); neutron porosity (NPHI); the photoelectric

index (PEF); potassium (POTA); uranium (URAN); thorium (THOR); thorium/potassium ratio (TKRT); thorium/uranium ratio (TURT); uranium/potassium ratio (UKRT); bulk formation density (RHOB); gamma ray (GR); deep resistivity (HDRS); and shallow resistivity (HMRS). MLP and MELM algorithms are optimized by PSO and GA optimizers. The MELM-GA/PSO models are used in one step, followed by another to determine the number of hidden layers and neurons in the network, and finally to identify the optimal weights and biases to apply to those layers and neurons. This leads to the combined GA-MELM-GA and PSO-MELM-PSO prediction models. After comparing the models used in this study, it was found that the MELM-PSO model has higher performance accuracy for data related to wells AJ#A, AJ#B and performance accuracy for this model for test data $RMSE = 0.0047$ 1/m; $R^2 = 0.9931$. The data field related to AJ # C was used to develop this model in comparison to other wells, and acceptable results with high accuracy were obtained with $R^2 = 0.9964$; $RMSE = 0.0041$ 1/m. According to the RMSE diagram vs iteration, it is clearly showing that the performance accuracy of the models is $MLP-PSO < MLP-GA < MELM-GA < MELM-PSO$, respectively. This method can be used in other fields, but it must be recalibrated with at least one

Table 12 FVDC prediction accuracy for the training subset in Aghajari oil field (AJ#A and AJ#B) (1463 Data Records; ~70%) based on 12-input variable feature selection

Models	APD	AAPD	SD	MSE	RMSE	R^2
Units	(%)	(%)	(1/m)	(1/m)	(1/m)	–
MLP-GA	0.030	0.463	0.048	2.288E-03	0.0478	0.9547
MLP-PSO	0.004	0.542	0.053	2.848E-03	0.0534	0.9322
MELM-GA	–0.004	0.108	0.011	1.216E-04	0.0110	0.9765
MELM-PSO	0.002	0.050	0.005	2.783E-05	0.0053	0.9903

Table 13 FVDC prediction accuracy for the testing subset in Aghajari oil field (AJ#A and AJ#B) (310 Data Records; ~15%) based on 12-input variable feature selection

Models	APD	AAPD	SD	MSE	RMSE	R^2
Units	(%)	(%)	(1/m)	(1/m)	(1/m)	–
MLP-GA	–0.086	0.341	0.032	1.032E-03	0.0321	0.9621
MLP-PSO	–0.080	0.390	0.036	1.275E-03	0.0357	0.9493
MELM-GA	–0.021	0.069	0.007	4.425E-05	0.0067	0.9842
MELM-PSO	–0.003	0.043	0.004	1.580E-05	0.0040	0.9959

Table 14 FVDC prediction accuracy for the validation subset in Aghajari oil field (AJ#A and AJ#B) (310 Data Records; ~15%) based on 12-input variable feature selection

Models	APD	AAPD	SD	MSE	RMSE	R^2
Units	(%)	(%)	(1/m)	(1/m)	(1/m)	–
MLP-GA	–0.115	0.792	0.043	1.831E-03	0.0428	0.9584
MLP-PSO	0.002	0.972	0.056	3.082E-03	0.0555	0.9408
MELM-GA	–0.005	0.190	0.011	1.276E-04	0.0113	0.9804
MELM-PSO	0.012	0.084	0.005	2.191E-05	0.0047	0.9931

Fig. 10 Measured versus predicted FVDC for each data record in the training, testing and total subset evaluated for four hybrid machine learning algorithms (MLP-GA, MLP-PSO, MELM-GA and MELM-

PSO) from the 2803 data point in the Aghajari oil field (AJ#A and AJ#B) based on 12-input variable feature selection

Fig. 11 FVDC prediction error histograms display with normal distributions (red line) for four hybrid machine learning algorithms (MLP-GA, MLP-PSO, MELM-GA and MELM-PSO) based on 2083 subset

data records from the Aghajari oil fields (AJ#A and AJ#B) based on 12-input variable feature selection

Fig. 12 FVDC's RMSE vs iteration for the training subset (drawn from wells AJ#A and AJ#B) for the four hybrid machine learning algorithms model (MLP-GA, MLP-PSO, MELM-GA and MELM-PSO) during supervised learning

Fig. 13 Predicted versus measured FVDC values compared for the MELM-PSO mode trained with data from well AJ#A and AJ#B applied to AJ#C data points

Table 15 FVDC prediction accuracy in term of six statistical metrics applying MELM-PSO model trained with AJ#A and AJ#B wells data records to datasets from other Aghajari oil well based on 12-input variable feature selection

Models	APD	AAPD	SD	MSE	RMSE	R^2
Units	(%)	(%)	(1/m)	(1/m)	(1/m)	–
MELM-PSO	–0.006	0.513	0.004	1.678E-05	0.0041	0.9964

well from another field before it can be used in another field. The developed method provides insights for the use of machine learning to improve accuracy, reduce errors, and avoid data overfitting in order to create the best possible prediction performance for FVDC prediction.

Acknowledgements This research was supported by Tomsk Polytechnic University development program.

Funding This research was supported by Tomsk Polytechnic University development program.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdali MR et al (2021) Petroleum well blowouts as a threat to drilling operation and wellbore sustainability: causes prevention safety and emergency response. *J Constr Mater/Special Issue Sustain Pet Eng* ISSN 2652:3752
- Aguilera R (2008) Role of natural fractures and slot porosity on tight gas sands, SPE Unconventional Reservoirs Conference. Society of Petroleum Engineers, Moscow
- AlAwad MN, Fattah K (2017) Utilization of shredded waste car tyres as a fracture seal material (FSM) in oil and gas drilling operations. *J Pet Gas Eng* 8:21–28. <https://doi.org/10.5897/JPGGE2017.0267>
- Ali, J. (1994). Neural networks: a new tool for the petroleum industry?, European petroleum computer conference. Society of Petroleum Engineers
- Allan J, Sun SQ (2003) Controls on recovery factor in fractured reservoirs: lessons learned from 100 fractured fields, SPE annual technical conference and exhibition. Society of Petroleum Engineers, Moscow
- Andersen MA et al (2013) Core truth in formation evaluation. *Oilfield Review* 25:16–25
- Atashnezhad A et al (2014) Designing and optimizing deviated wellbore trajectories using novel particle swarm algorithms. *J Nat Gas Sci Eng* 21:1184–1204. <https://doi.org/10.1016/j.jngse.2014.05.029>
- Bhattacharya S, Mishra S (2018) Applications of machine learning for facies and fracture prediction using Bayesian network theory and random forest: case studies from the Appalachian basin, USA. *J Petrol Sci Eng* 170:1005–1017. <https://doi.org/10.1016/j.petrol.2018.06.075>
- Boadu FK (1998) Inversion of fracture density from field seismic velocities using artificial neural networks. *Geophysics* 63:534–545. <https://doi.org/10.1190/1.1444354>
- Chandrashekar G, Sahin F (2014) A survey on feature selection methods. *Comput Electr Eng* 40:16–28. <https://doi.org/10.1016/j.compeleceng.2013.11.024>
- Cheng J, Xiong Y (2017) Application of extreme learning machine combination model for dam displacement prediction. *Proced Comput Sci* 107:373–378. <https://doi.org/10.1016/j.procs.2017.03.120>
- Choubineh A et al (2017) Improved predictions of wellhead choke liquid critical-flow rates: modelling based on hybrid neural network training learning based optimization. *Fuel* 207:547–560. <https://doi.org/10.1016/j.fuel.2017.06.131>
- Darvishpour A et al (2019) Wellbore stability analysis to determine the safe mud weight window for sandstone layers. *Pet Explor Dev* 46:1031–1038. [https://doi.org/10.1016/S1876-3804\(19\)60260-0](https://doi.org/10.1016/S1876-3804(19)60260-0)
- Dutta P et al (2007) A novel approach to fracture characterization utilizing borehole seismic data, SPE Middle East Oil and Gas Show and Conference. OnePetro. <https://doi.org/10.2118/105427-MS>
- Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory, MHS'95. Proceedings of the 6th international symposium on micro machine and human science. Ieee, pp 39–43. <https://doi.org/10.1109/MHS.1995.494215>
- Farsi M et al (2021) Predicting formation pore-pressure from well-log data with hybrid machine-learning optimization algorithms. *Nat Resour Res* 30(5):3455–3481
- Feng Y et al (2016) A review on fracture-initiation and-propagation pressures for lost circulation and wellbore strengthening. *SPE Drill Complet* 31:134–144. <https://doi.org/10.2118/181747-PA>
- Gale JF et al (2007) Natural fractures in the Barnett Shale and their importance for hydraulic fracture treatments. *AAPG Bull* 91:603–622. <https://doi.org/10.1306/110106060601>
- Ghasemi M, Bayuk I (2020) Bounds for pore space parameters of petroelastic models of carbonate rocks. *Izvestiya, Phys Solid Earth* 56(2):207–224
- Ghorbani H, Moghadasi J (2014) Development of a new comprehensive model for choke performance correlation in iranian oil wells. *Adv Environ Biol* 8:877–882
- Ghorbani H et al (2014) Development of a new comprehensive model for choke performance correlation in iranian gas condensate wells. *Adv Environ Biol* 8:308–313

- Ghorbani H et al (2017) The exposure of new estimating models for bubble point pressure in crude oil of one of the oil fields in Iran. *Am J Oil Chem Technol*. <https://doi.org/10.1080/10916466.2020.1867168>
- Ghorbani H et al (2017) Developing a new multiphase model for choke function relation for Iran's gas wells. *Am J Oil Chem Technol* 5:172–180
- Ghorbani H et al (2018) Prediction of oil flow rate through an orifice flow meter: artificial intelligence alternatives compared. *Petroleum*. <https://doi.org/10.1016/j.petlm.2018.09.003>
- Ghorbani H et al (2020) Adaptive neuro-fuzzy algorithm applied to predict and control multi-phase flow rates through wellhead chokes. *Flow Meas Instrum* 76:101849. <https://doi.org/10.1016/j.flowmeasinst.2020.101849>
- Hassanpouryouzband A et al (2018a) CO₂ capture by injection of flue gas or CO₂–N₂ mixtures into hydrate reservoirs: dependence of CO₂ capture efficiency on gas hydrate reservoir conditions. *Environ Sci Technol* 52:4324–4330. <https://doi.org/10.1021/acs.est.7b05784>
- Hassanpouryouzband A et al (2018b) Insights into CO₂ capture by flue gas hydrate formation: gas composition evolution in systems containing gas hydrates and gas mixtures at stable pressures. *ACS Sustain Chem Eng* 6:5732–5736. <https://doi.org/10.1021/acssuschemeng.8b00409>
- Hassanpouryouzband A et al (2019) Geological CO₂ capture and storage with flue gas hydrate formation in frozen and unfrozen sediments: method development, real time-scale kinetic characteristics, efficiency, and clathrate structural transition. *ACS Sustain Chem Eng* 7:5338–5345. <https://doi.org/10.1021/acssuschemeng.8b06374>
- Hazbeh O et al (2021) Comparison of accuracy and computational performance between the machine learning algorithms for rate of penetration in directional drilling well. *Pet Res*. <https://doi.org/10.1016/j.ptlrs.2021.02.004>
- Hazbeh O et al (2021b) Hybrid computing models to predict oil formation volume factor using multilayer perceptron algorithm. *J Pet Min Eng*. <https://doi.org/10.21608/jpme.2021.52149.1062>
- Huang G-B et al (2011) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern Part B* 42:513–529. <https://doi.org/10.1109/TSMCB.2011.2168604>
- Ince R (2004) Prediction of fracture parameters of concrete by artificial neural networks. *Eng Fract Mech* 71:2143–2159. <https://doi.org/10.1016/j.engfracmech.2003.12.004>
- Ja'fari A et al (2012) Fracture density estimation from petrophysical log data using the adaptive neuro-fuzzy inference system. *J Geophys Eng* 9:105–114. <https://doi.org/10.1088/1742-2132/9/1/013>
- Jain A, Zongker D (1997) Feature selection: evaluation, application, and small sample performance. *IEEE Trans Pattern Anal Mach Intell* 19:153–158. <https://doi.org/10.1109/34.574797>
- John GH et al (1994) Irrelevant features and the subset selection problem. *Machine learning proceedings 1994*. Elsevier, Amsterdam, pp 121–129
- Khorzoughi MB et al (2018) Rock fracture density characterization using measurement while drilling (MWD) techniques. *Int J Min Sci Technol* 28:859–864. <https://doi.org/10.1016/j.ijmst.2018.01.001>
- Kuramoto SI et al (2008) The first D/V Chikyu IODP operations: successful logging and coring during NanTroSEIZE stage 1 expeditions. *Sci Drill* 6:38–42
- Li T et al (2018) Prediction of fracture density using genetic algorithm support vector machine based on acoustic logging data. *Geophysics* 83:D49–D60. <https://doi.org/10.1190/geo2017-0229.1>
- Liang N-Y et al (2006) A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Trans Neural Netw* 17:1411–1423. <https://doi.org/10.1109/TNN.2006.880583>
- Luthi SM (2001) Electrical borehole imaging. *Geological Well Logs*. Springer, Heidelberg, pp 74–123
- Mirjalili S (2019) Genetic algorithm, Evolutionary algorithms and neural networks. Springer, Heidelberg, pp 43–55
- Mohammadian N, Ghorbani H (2015) An investigation on chemical formation damage in Iranian reservoir by focus on mineralogy role in shale swelling potential in Pabdeh and Gurpi formations. *Adv Environ Biol* 9:161–166
- Mohamadian N et al (2018) Rheological and filtration characteristics of drilling fluids enhanced by nanoparticles with selected additives: an experimental study. *Adv Geo-Energy Res* 2:228–236
- Mohamadian N et al (2019) A hybrid nanocomposite of poly (styrene-methyl methacrylate-acrylic acid)/clay as a novel rheology-improvement additive for drilling fluids. *J Polym Res*. <https://doi.org/10.1007/s10965-019-1696-6>
- Mohamadian N et al (2021) A geomechanical approach to casing collapse prediction in oil and gas wells aided by machine learning. *J Petrol Sci Eng* 196:107811. <https://doi.org/10.1016/j.petrol.2020.107811>
- Murphy DP (1993). What's new in MWD and formation evaluation. [Measurement while drilling (MWD)]. *World Oil* 214. <https://www.osti.gov/biblio/6512488>.
- Nelson R (1985) Geologic analysis of fractured reservoirs: contributions in petroleum geology and engineering. Gulf Publishing Company, Houston
- Nelson R (2001) Geologic analysis of naturally fractured reservoirs. Elsevier, Amsterdam
- Nouri-Taleghani M et al (2015) Fracture density determination using a novel hybrid computational scheme: a case study on an Iranian Marun oil field reservoir. *J Geophys Eng* 12:188–198. <https://doi.org/10.1088/1742-2132/12/2/188>
- Ouenes A (2000) Practical application of fuzzy logic and neural networks to fractured reservoir characterization. *Comput Geosci* 26:953–962. [https://doi.org/10.1016/S0098-3004\(00\)00031-5](https://doi.org/10.1016/S0098-3004(00)00031-5)
- Rashidi S et al (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. <https://doi.org/10.1016/j.jappgeo.2020.104205>
- Rashidi S et al (2021) Determination of bubble point pressure & oil formation volume factor of crude oils applying multiple hidden layers extreme learning machine algorithms. *J Petrol Sci Eng*. <https://doi.org/10.1016/j.petrol.2021.108425>
- Romano C et al (2019) Automated high accuracy, rapid beam hardening correction in X-ray computed tomography of multi-mineral, heterogeneous core samples. *Comput Geosci* 131:144–157. <https://doi.org/10.1016/j.cageo.2019.06.009>
- Sabah M et al (2019) A machine learning approach to predict drilling rate using petrophysical and mud logging data. *Earth Sci Inform* 12:319–339. <https://doi.org/10.1007/s12145-019-00381-4>
- Sarkheil H et al (2009) The fracture network modeling in naturally fractured reservoirs using artificial neural network based on image logs and core measurements. *Aust J Basic Appl Sci* 3:3297–3306
- Sarkheil H et al (2013) Fractures distribution modeling using fractal and multi-fractal–neural network analysis in Tabnak hydrocarbon field Fars Iran. *Arabian J Geosci* 6:945–956
- Serra O (1989) Formation MicroScanner image interpretation. Schlumberger Educational Services, Houston
- Simon D (2013) Evolutionary optimization algorithms. Wiley, Hoboken
- Su X et al (2019) Prediction model of hot metal temperature for blast furnace based on improved multi-layer extreme learning machine. *Int J Mach Learn Cybern* 10:2739–2752. <https://doi.org/10.1007/s13042-018-0897-3>
- Suboyin A et al (2020) Hydraulic fracturing design considerations, water management challenges and insights for Middle Eastern

- shale gas reservoirs. *Energy Rep* 6:745–760. <https://doi.org/10.1016/j.egy.2020.03.017>
- Tang, J., et al. (2014). A fast learning algorithm for multi-layer extreme learning machine, 2014 IEEE international conference on image processing (ICIP). IEEE, pp 175–178. <https://doi.org/10.1109/ICIP.2014.7025034>.
- Thompson L (2000) Fractured reservoirs: integration is the key to optimization. *J Petrol Technol* 52:52–54. <https://doi.org/10.2118/56010-JPT>
- Tokhmechi B et al (2009) Fracture detection from water saturation log data using a Fourier–wavelet approach. *J Petrol Sci Eng* 69:129–138. <https://doi.org/10.1016/j.petrol.2009.08.005>
- Toyobo O et al (2020) Benefits of real-time LWD static borehole image in reservoir navigation service. SPE Nigeria annual international conference and exhibition. Society of Petroleum Engineers. <https://doi.org/10.2118/203717-MS>.
- Wahab MNA et al (2015) A comprehensive review of swarm optimization algorithms. *PLoS One* 10:1–36. <https://doi.org/10.1371/journal.pone.0122827>
- Wang H, Sharma MM (2019) Determine in-situ stress and characterize complex fractures in naturally fractured reservoirs from diagnostic fracture injection tests. *Rock Mech Rock Eng* 52:5025–5045. <https://doi.org/10.1007/s00603-019-01793-w>
- Warpinski NR et al (2009) Stimulating unconventional reservoirs: maximizing network growth while optimizing fracture conductivity. *J Can Pet Technol* 48:39–51. <https://doi.org/10.2118/114173-PA>
- Xu L et al (2016) Spectral gamma-ray measurement while drilling. *Petrophys SPWLA J Form Eval Reser Descr* 57:377–389
- Xu C et al (2019) Friction coefficient: a significant parameter for lost circulation control and material selection in naturally fractured reservoir. *Energy* 174:1012–1025. <https://doi.org/10.1016/j.energy.2019.03.017>
- Yeom C-U, Kwak K-C (2017) Short-term electricity-load forecasting using a TSK-based extreme learning machine with knowledge representation. *Energies* 10:1613. <https://doi.org/10.3390/en10101613>
- Zazoun RS (2013) Fracture density estimation from core and conventional well logs data using artificial neural networks: the Cambro-Ordovician reservoir of Mesdar oil field, Algeria. *J Afr Earth Sc* 83:55–73. <https://doi.org/10.1016/j.jafrearsci.2013.03.003>
- Zellou A, Ouenes A (2003) Integrated fractured reservoir characterization using neural networks and fuzzy logic: Three case studies. *Developments in petroleum science*. Elsevier, Amsterdam, pp 583–602
- Zerrouki AA et al (2014) Prediction of natural fracture porosity from well log data by means of fuzzy ranking and an artificial neural network in Hassi Messaoud oil field, Algeria. *J Petrol Sci Eng* 115:78–89. <https://doi.org/10.1016/j.petrol.2014.01.011>
- Zhang N et al (2016) Denoising Laplacian multi-layer extreme learning machine. *Neurocomputing* 171:1066–1074. <https://doi.org/10.1016/j.neucom.2015.07.058>
- Zheng L et al (2019) Research of bearing fault diagnosis method based on multi-layer extreme learning machine optimized by novel ant lion algorithm. *IEEE Access* 7:89845–89856. <https://doi.org/10.1109/ACCESS.2019.2926348>

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