Fuel

Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields

--Manuscript Draft--

Manuscript Number:	JFUE-D-21-01682R3
Article Type:	Research Paper
Keywords:	Gas flow rate, multi-hidden layer extreme learning machine, hybrid machine learning algorithms; least squares support vector machine, wellhead choke.
Corresponding Author:	Hamzeh Ghorbani Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran IRAN, ISLAMIC REPUBLIC OF
First Author:	Abouzar Rajabi Behesht Abad
Order of Authors:	Abouzar Rajabi Behesht Abad
	Hamzeh Ghorbani
	Nima Mohamadian
	Shadfar Davoodi
	Mohammad Mehrad
	Saeed Khezerloo-ye Aghdam
	Hamid Reza Nasriani
Abstract:	Condensate reservoirs are the most challenging hydrocarbon reservoirs in the world. The behavior of condensate gas reservoirs regarding pressure and temperature variation is unique. Adjusting fluid flow rate through wellhead chokes of condensate gas wells is critical and challenging for reservoir management. Predicting this vital parameter is a big step for the development of condensate gas fields. In this study, a novel machine learning approach is developed to predict gas flow rate (Q_g) from six input variables: temperature (T); upstream pressure (P_u); downstream pressure (P_d); gas gravity (γ_g); choke diameter (D_{64}) and gas-liquid ratio (GLR). Due to the absence of accurate recombination methods for determining Q_g, machine learning methods offer a functional alternative approach. Four hybrid machine learning (HLM) algorithms are developed by integrating multiple extreme learning machine (MELM) and least squares support vector machine (LSSVM) with two optimization algorithms, the genetic algorithm (GA) and the particle swarm optimizer (PSO). The evaluation conducted on prediction performance and accuracy of the four HLM models developed indicates that the MELM-PSO model has the highest Q_g prediction accuracy achieving a root mean squared error (RMSE) of 2.8639 MScf/Day and a coefficient of determination (R^2) 0.9778 for a dataset of 1009 data records compiled from gas-condensate fields around Iran. Comparison of the prediction performance of the HLM models developed with those of the previous empirical equations and artificial intelligence models reveals that the novel MELM-PSO model presents superior prediction efficiency and higher computational accuracy. Moreover, the Spearman correlation coefficient analysis performed demonstrates that D_{64} and GLR are the most influential variables in the gas flow rate for the large dataset evaluated in this study.

Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields

Highlights

- 1009 record date of from the Iranian condensate fields (Marun-Khami, Aghajari-Khami and Ahvaz-Khami).
- New hybrid machine learning technique accurately predicts gas flow rate through wellhead choke in gas condensate reservoirs.
- MELM-PSO model constructs the most accurate condensate gas flow rate predictions.
- Choke size (D_{64}), downstream pressure (P_d) and gas liquid ratio (GLR) have the greatest influence.

Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields

By

Abouzar Rajabi Behesht Abad
Department of Petroleum Engineering, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran
drrajabi.iput@yahoo.com

Hamzeh Ghorbani [Corresponding author]
Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
hamzehghorbani68@yahoo.com
orcid.org/0000-0003-4657-8249

Nima Mohamadian
Young Researchers and Elite Club, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran
nima.0691@gmail.com

Shadfar Davoodi
School of Earth Sciences & Engineering, Tomsk Polytechnic University, Lenin Avenue, Tomsk, Russia
davoodis@hw.tpu.ru
orcid.org/0000-0003-1733-1677

Mohammad Mehrad
Faculty of Mining, Petroleum and Geophysics Engineering, Shahrood University of Technology, Shahrood, Iran
mmehrad1986@gmail.com

Saeed Khezerloo-ye Aghdam
Department of petroleum engineering, Amirkabir University of Technology, Tehran, Iran
saeed.khezerloo@gmail.com

Hamid Reza Nasriani
School of Engineering, Faculty of Science and Technology, University of Central Lancashire, Preston, United Kingdom
hrnasriani@uclan.ac.uk
orcid.org/0000-0001-9556-7218

Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields

Abstract

Condensate reservoirs are the most challenging hydrocarbon reservoirs in the world. The behavior of condensate gas reservoirs regarding pressure and temperature variation is unique. Adjusting fluid flow rate through wellhead chokes of condensate gas wells is critical and challenging for reservoir management. Predicting this vital parameter is a big step for the development of condensate gas fields. In this study, a novel machine learning approach is developed to predict gas flow rate (Q_g) from six input variables: temperature (T); upstream pressure (P_u); downstream pressure (P_d); gas gravity (γ_g); choke diameter (D_{64}) and gas-liquid ratio (GLR). Due to the absence of accurate recombination methods for determining Q_g, machine learning methods offer a functional alternative approach. Four hybrid machine learning (HLM) algorithms are developed by integrating multiple extreme learning machine (MELM) and least squares support vector machine (LSSVM) with two optimization algorithms, the genetic algorithm (GA) and the particle swarm optimizer (PSO). The evaluation conducted on prediction performance and accuracy of the four HLM models developed indicates that the MELM-PSO model has the highest Q_g prediction accuracy achieving a root mean squared error (RMSE) of 2.8639 MScf/Day and a coefficient of determination (R^2) 0.9778 for a dataset of 1009 data records compiled from gas-condensate fields around Iran. Comparison of the prediction performance of the HLM models developed with those of the previous empirical equations and artificial intelligence models reveals that the novel MELM-PSO model presents superior

prediction efficiency and higher computational accuracy. Moreover, the Spearman correlation coefficient analysis performed demonstrates that D_{64} and GLR are the most influential variables in the gas flow rate for the large dataset evaluated in this study.

Keywords: Gas flow rate, multi-hidden layer extreme learning machine, hybrid machine learning algorithms; least squares support vector machine, wellhead choke.

1. Introduction

Hydrocarbon fuels are still recognized worldwide as the driving force and strategic energy to develop leading economic and industrial goals [1-3]. A sustainable production approach from hydrocarbon reservoirs is an essential production management policy that enables upstream companies to exploit hydrocarbon reservoirs efficiently [4]. Regardless of the economic perspective, controlling the production rate by wellhead chokes is the most important management lever for optimizing the production process. Increasing the production rate without involving engineering concerns adversely affects wells' productivity and shortens their production life [5]. Such problems will be exacerbated, especially in unconventional gas reservoirs with tight carbonated structure and very low permeability [6]. The unique phase behavior of condensate gas makes the production rate control techniques even more challenging and vital in such reservoirs. [7]. In condensate reservoirs, the production rate declines significantly due to the accumulation of unproducible liquid in the near-wellbore region [8]. The reservoir fluid in the regions far from the wellbore is a combination of rich gas and non-moveable connate water. At the early production stage, the pressure drops below the dew point near the wellbore region, and the rich gas is converted into condensate. This isothermal condensation is known as retrograde condensation [9]. The accumulation of valuable condensate droplets around the wellbore, also known as the condensate bank/ring, has not yet reached critical saturation for portability, resulting in a positive skin factor [10]. Production from gas condensate reservoirs requires meticulous planning and management [11]. Scheduled production plans for sale and export contracts of gas and gas condensate productive [12] require continuous production at the desired rate. Any disruption to the production process may damage economic obligations. Therefore, accurate control and management of production rates and pressure drop through production wells are essential to implement sustainable production programs from condensate reservoirs. By understanding the importance of preserving and sustainable production from gas condensate resources, the position and credibility of efficient tools for control and handling of this vital goal become clearer. Wellhead chokes are a very cost-effective and efficient tool for measuring and controlling multiphase flow rates at an optimum level [13]. Accurate measurement of multiphase flow is one of the concerns of production engineers [14]. The values determined in these measurements are the basic input parameters for calculation in many reservoir performance relationships. Determination of multiphase flow rate is crucial in planning and adopting correct measures and reforms in production policies commensurate with the reservoir's performance during operation [15]. The back pressure applying by wellhead chokes has several advantages, such as stabilizing the multiphase flow rate [16], preventing further pressure drop at the bottom hole section and condensate drop out, avoiding to create the skin factor due to pressure drop, and preventing water coning in gas condensate reservoirs [17, 18]. Numerous experimental and theoretical relationships have been introduced to estimate the multiphase flow rate through wellhead chokes. In most of them, the basis of flow calculations depends on the pressure difference between the upstream and downstream instruments [19-21]. One of the most popular computational models proposed belongs to Gilbert (1954), which has been widely used to calculate the liquids flow rate through the wellhead choke and in recent years has been adapted for data from different regions (show in Eq. (1)) [22-26]:

$$Q_{liq} = M \frac{P_{wh} D_{64}^{O}}{GLR^{l}} \tag{1}$$

Where Q_{liq} is the rate of liquids production (STB/D), P_{wh} is the wellhead pressure (psi), D_{64} is the choke size (1/64 inch), GLR is the gas to liquid ratio (SCF/STB), and M, l, O are experimental coefficients calculated where sufficient data is available for specific reservoir systems.

Osman and Dokla 1990 used a dataset from gas condensate wells in the Middle East region to develop an empirical relationship for calculating the flow through the wellhead chokes [27]. They adapted the Gilbert equation in three modified forms by changing the pressure parameters (replacing the upstream pressure with the pressure drop across the choke) for the wells' data in gas condensate reservoirs. Guo et al. 2002 evaluated data from 239 condensate gas wells with Sachdeva's multiphase choke flow equation and compared the results with field measurements. After receiving the under-estimated performance feedback from this model, they could adapt it using different choke discharge coefficients (CD) to obtain less computational error [28]. Al-Attar 2008 developed an empirical equation to describe a sub-critical flow model in gas condensate wellhead chokes ranging from 24/64 to 128/64 inches for different choke sizes [29]. Nasriani and Kalantariasl (2019) also presented a tuned equation derived from the Gilbert basic equation to measure flow rate in sub-critical flow regime based on data collected from 50 wells in some gas condensate reservoirs in southern Iran [30]. Seidi and Sayahi (2015), by adapting Gilbert's basic equation using the genetic algorithm and nonlinear regression methods and applying them to 67 datasets gathered from different gas condensate fields, proposed an optimized model for estimating the condensate gas flow rate [31]. The equations presented by these researchers are summarized in Table 1.

Recently, some researchers strived to solve many oil, gas and geological hydrogen storage [32-36]. However, data science has provided a new way to move from conventional computing systems to faster, more accurate, and cost-effective computing methods. Today, new machine learning techniques are efficient tools for optimization and sophisticated computing that reduce operating costs and improve system performance. Extensive research has been conducted in recent years on the application of intelligent machine learning methods in various sectors of the upstream oil and gas industry, such as desalting system analysis [37], hydrocarbon phase behavior prediction [38-42], determination of oil and gas flow through orifice [43-46] and determination of flow rate through wellhead choke [18, 47-53]. Predicting multiphase flow rate from wellhead chokes is the subject of other studies on machine learning application in flow measurement concepts. Table 2 summarizes some of the recently published research on these smart models' performance in this field.

As shown in Table 2, in recent years, intelligent machine learning models for accurately estimating the flow rate of hydrocarbon fluids passing through wellhead chokes have been inexpensive, fast, and accurate solutions for calculating the production flow of hydrocarbon fluids. Machine learning models require a large and extensive range of data set to create a comprehensive and more accurate model. There is still a shortage of model construction by vast data sets specifically structured to predict gas flow rates.

Table 1 provides a comparison of previous empirical relationships, and Table 2 shows the results of intelligent methods proposed in previous studies. It is worth noting that the methods proposed in this paper are compared with those empirical methods in previous studies that presented better performance. In addition, as shown in Table 2, a limited number of studies have been performed on the gas flow rate prediction in gas & gas condensate reservoirs. As a result, this research, based on a database made of more than 1009 data records, has endeavored to develop novel models for gas flow rate prediction (MELM with PSO/GA optimizer) with minimized RMSE. The model developed employs six input variables, including temperature (T), upstream pressure (P_u), downstream pressure (P_d), gas gravity (γ_g), choke diameter (D_{64}), and gas-liquid ratio (GLR) to accurately predict gas flow rate from wellhead chock. Moreover, to create the best possible prediction performance and accuracy as well as to avoid overfitting, several control measures are applied in the present study.

Table 1. Empirical equations proposed by some researchers to determine the flow rate of condensate gas through wells.

Year	Authors / Reference	Formula	Dataset	Units	Coefficient	R^2	Error Functions
1990	Osman & Dokla [27]	$Q_g = a * \frac{{P_u}^b * {D_{64}}^c}{LGR^d}$	87 data points	Q_g: MScf/Day, P_u: Psia, D_{64}: inch, LGR: STB/MScf	a= 0.00130, b=1, c= 1.8298, d= 0.5598	-	Best result: AAPD%= 10.64
2008	Al-Attar [29]	$Q_g = a * \Delta p^b * {D_{64}}^c * GLR^d$	97 data points	Q_g: MMScf/Day, Δp: Psi, D_{64}: inch, LGR: STB/MScf	a= 3.37230e-5, b=1, c= 1.15537, d= 0.84695	Best result: 0.9521	Best result: AAPD%= 7.144
2015	Seidi and Sayahi [31]	$Q_g = a * \frac{\Delta p^b * {D_{64}}^c}{LGR^d}$	106 data points	Q_g: MMScf/Day, Δp: Psi, D_{64}: inch, LGR: STB/MMScf	a= 0.015, b=0.65, c= 1.27, d= 0.4	Best result: 0.9161	Best result: APD%= 23.93
2017	Ghorbani et al. [18]	$Q_g = aD_{64}{}^b \left(\frac{P_u}{14.7}\right)^c \sqrt{\left(\frac{1}{\gamma_g T}\right)^d \left[\left(\frac{P_d}{P_u}\right)^e - \left(\frac{P_d}{P_u}\right)^f\right]}$	92 data points	Q_g: MScf/Day, P_u and P_d: Psig, D_{64}: inch, γ_g: -, T: ^0F	a= 0.0001, b= 2.3481935, c= 1, d= 0.0001, e= 1:0360972, f= 1.498291	0.9677	APD%= 5.32

2019	Nasriani et al. [30]	$Q_g = a * \frac{\Delta p^b * D_{64}{}^c}{LGR^d}$	234 data points	Q_g: MMScf/Day, Δp: Psi, D_{64}: inch, LGR: STB/MMScf	a= 0.0437, b=0.4836, c= 1.1136, d= 0.3129	Best result: 0.97	Best result: AAPD%= 8.71

Table 2. Implementation of some machine learning algorithms to predict oil, gas, and gas condensate flow rates through wellhead wells.

Fluid Flow Type	Authors / Year	Machine Learning Techniques	Dataset	Input Parameters	R^2	Error Functions
Oil flow rate	Payaman & Salavati (2012) [54]	Artificial Neural Network (ANN)	196 data points	P_u - D_{64} - GOR	0.98	APD%= -0.33
	Nejatian et, al (2014) [55]	Least-Squares Support Vector Machine (LSSVM)	171 data point	Reynolds number - d/D - Choke flow coefficient	0.99	AAPD%= 0.256
	Gholgheysari Gorjaei et, al. (2015) [56]	Particle swarm optimization (PSO)-Least square support vector machine (LSSVM -PSO)	276 data points	P_u - D_{64} - GLR	0.965	APD%= -0.80

	Rostami & Ebadi (2017) [57]	Gene expression programming (GEP)	119 data points	P_u - D_{64} - GOR - γ_g - API	0.96	AAPD%= 14.808
	Ghorbani et, al. (2019) [50]	Genetic Algorithm and solver optimizers	127 data points	P_u - D_{64} - GLR - BS&W%	0.99	AAPD%=7.33
	Ghorbani et, al. (2020) [49]	Adaptive Neuro Fuzzy Inference System (ANFIS)	182 data points	P_u - D_{64} - GLR - BS&W%	0.998	AAPD%= 6.62
Oil flow rate assisted with gas lift	Khan et al. (2020) [51]	ANN	1950 data points	P_u - D_{64} - Tup - Pd - Oil API	0.99	AAPD%= 2.56
Gas flow rate in gas & gas condensate reservoir	ZareNezhad & Aminian (2011) [58]	ANN	97 data points	ΔP - GOR - D_{64}	0.99	APD%= 0.486
	Elhaj et, al. (2015) [59]	ANN	162 data points	P_u - D_{64} - P_d - T- γ_g	0.99	AAPD%= 0.828
		Fuzzy Logic (FL)			0.97	AAPD%= 0.681
	Kalam et, al. (2019) [59]	ANN	17097 data points	P_u - D_{64}- T – Q_g	0.953	AAPD%= 7.386
		Functional Network (FN)			0.91	AAPD%= 12
		ANFIS			0.95	AAPD%= 14

2. Methodology

2.1. Work Flow

A systematic methodology involving ten steps (Fig. 1) is developed for constructing and evaluating the four hybrid machine learning algorithms employed for the prediction of gas flow rate through wellhead chokes. The first step in the proposed workflow is data gathering from gas condensate fields. Next, the maximum and minimum values of variables need to be determined. Afterward, the variables are normalized between -1 and +1 (Eq. (2)). Once the data are normalized, the set of data is divided into two subsets, training and testing. Then, the machine learning optimizer's accuracy is determined by statistical indicators such as AAPD%, SD, MSE, RMSE, and R^2. Results obtained from accuracy evaluation are compared with empirical equations and hybrid machine learning techniques [47].

$$x_i^l = \left(\frac{x_i^l - xmin^l}{xmax^l - xmin^l}\right) * 2 - 1 \tag{2}$$

Where x_i^l is the value of attribute l for data record *I*; $xmin^l$ is the minimum value of the attribute l among all the data records in the dataset; and $xmax^l$ is the maximum value of the attribute l among all the data records in the dataset.

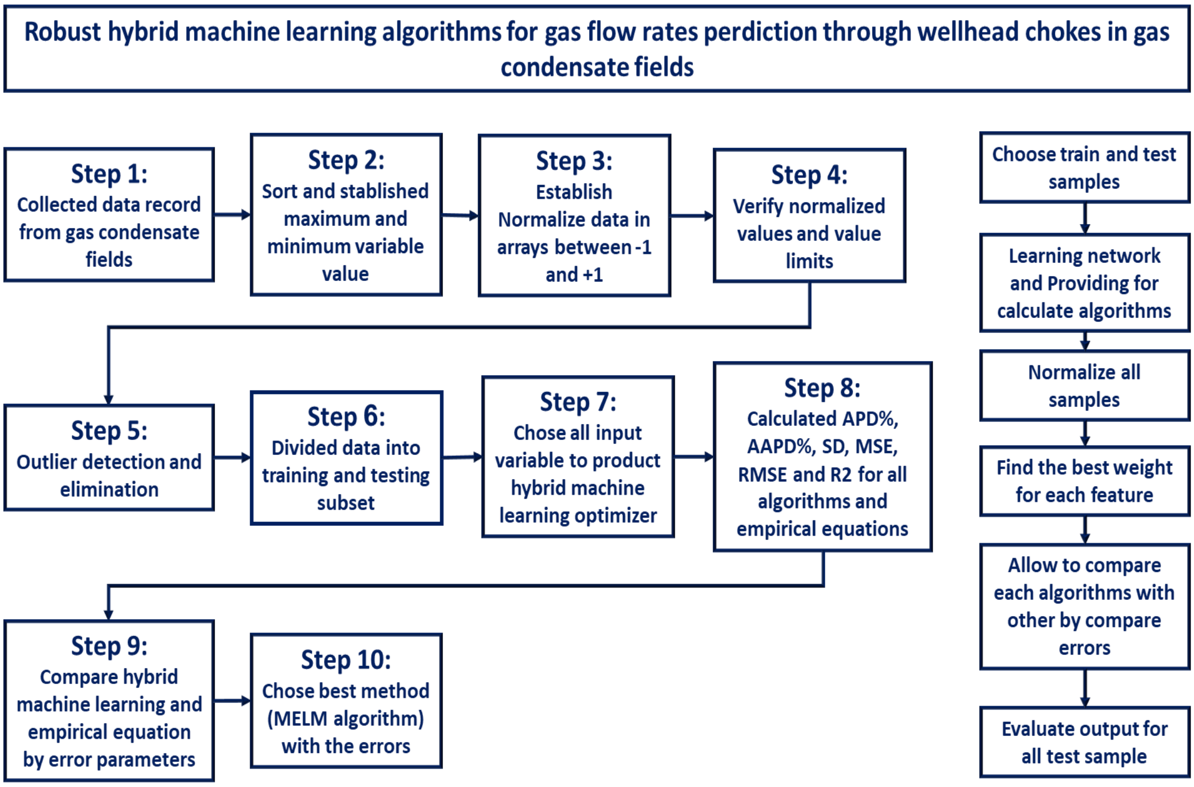

Fig. 1. Schematic of workflow proposed for constrction and evaluation of four HLM algorithms used for Q_g prediction.

2.2. Least square support vector machine (LSSVM)

The least-square support vector machine (LSSVM) is an expanded version of the support vector machine (SVM) that Suykens and Vandewalle developed in 1998 [60, 61]. LSSVM technique uses powerful features of SVM [62, 63]. However, there are two major differences between the LSSVM and SVM learning techniques. First, the LSSVM technique uses square errors in the cost function instead of nonnegative errors, and second, the LSSVM technique applies equality constraints instead of inequality constraints. Consequently, in LSSVM, a linear system of equations is solved instead of a quadratic programming problem, leading to a considerable reduction in the learning model's computational time [64, 65].

In the LSSVM method, the following nonlinear cost function (Eq. (3)) is used for approximation [66, 67]:

$$f(x) = w^{T}\phi(x_i) + b \quad (3)$$

In which x_i denotes the input variable to the function, the dimension of which is $N \times n$, where N and n stand for the number of samples in the dataset and the number of inputs parameters, respectively. w and b represent the weight and bias vector of output layer respectively, $\phi(x_i)$ indicates kernel function, T is transpose matrix. For the sake of brevity, the readers are advised to refer to the previously published works, where a detailed theoretical description of the LSSVM model is provided [61, 62, 68-73]. Since the LSSVM model parameters have a considerable influence on the model accuracy and performance, GA and PSO optimization algorithms were applied for optimizing those parameters in the present study. Besides these control parameters, the type of kernel applied in LSSVM model construction also has a pronouncing effect on the performance and accuracy of the LSSVM model. Given that there is no standard way in kernel function selection, four of the most commonly applied kernel functions, including the linear kernel, polynomial kernel, radial basis function kernel, and multilayer perceptron kernel, have been tested out in the present study. Among those, the RBF kernel is found to be the most efficient one.

2.3. Multilayer extreme learning machine (MELM)

The extreme learning machine (ELM), as a new quick single hidden layer feedforward network, was first developed by Huang et al. in 2005 [74]. Since its emergence, ELM has been widely used in generating solutions to various problems, namely regression, classification, and clustering. The basic structure of ELM resembles a single hidden layer backpropagation (BP) neural network that is composed of three layers which are input, hidden, and output layers. However, the method used in training ELM is soundly different from that of the conventional network. Indeed, the ELM technique randomly assigns the hidden parameters, the hidden nodes biases, and the input weights to hidden nodes and analytically calculates the output weights. As a result, the time required for optimizing the hidden parameters of the model is significantly decreased by avoiding iterative calculations during model training [75, 76]. Elaboration on structures and the theoretical principles of conventional artificial neural networks and ELM models can be discovered in previous publications [74, 77-79].

Complex variants of ELM with several hidden layers are recommended to solve problems with a nonlinear dataset of high complexity. Therefore, a complex form of ELM that includes multiple hidden layers, called MELM, was developed based on the deep learning (DL) concept [80]. The construction procedure of the MELM learning model is elaborated in recently published works [38, 63].

2.4. Optimization algorithm techniques

2.4.1. Genetic algorithm (GA)

Genetic algorithm is a class of evolutionary algorithms developed based on natural selection and evaluation principles. This method is commonly applied for solving search and optimization problems. This method obtains the global optimum solution within a complex multi-dimensional space. In the GA method, the poorer population of parents is replaced with the better offspring population by each generation of the population using three operations: selection, crossover, and mutation. This process is reaping by the GA until a high accuracy of prediction is achieved. Hence, the population's final output individual is the best parameter group [81, 82]. Fig. 2 illustrates the cycle of GA. To keep the study concise, the readers are advised to read previously published studies in which detailed theoretical descriptions on the GA technique are provided [83-86].

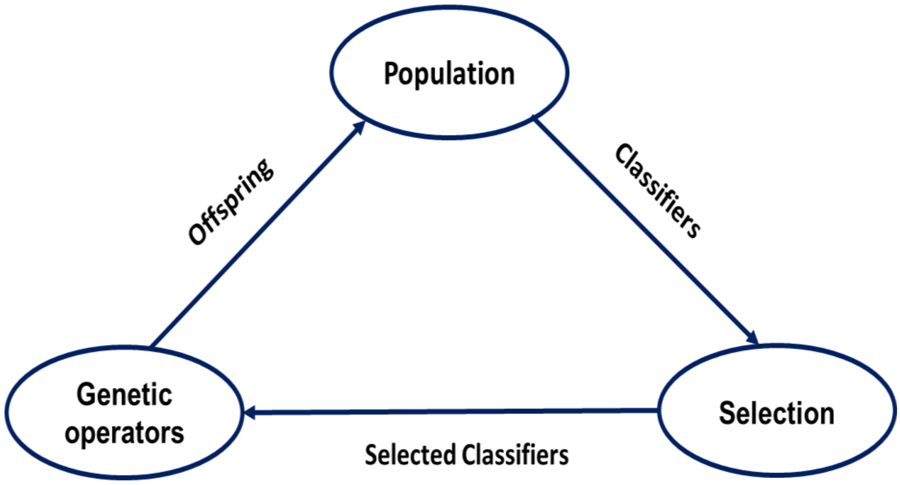

Fig. 2. Schematic of GA cycle.

2.4.2. Particle swarm optimization (PSO)

Particle Swarm Optimization (PSO), an optimization algorithm inspired by natural swarming and flocking of birds and insects, was proposed by Kennedy and Eberhart [87]. This optimization method initiates a population or "swarm" made of random solutions and, by updating generation, attempts to obtain the optimal solution. In the PSO algorithm, solutions are named "particles" [38]. The population particles go through the space of the problem by following the current best particles in the population. Each of the population particles possesses a velocity and a position, and they seek positions with good fitness in the space. During the optimization process, two main pieces of information are memorized by each particle i) the best position heaving been so far visited by the particle (Pb) ii) the global best position attained by the particles in the whole swarm (Gb) [29, 38]. To obtain the best solution, several iterations are performed by PSO. In each step, the solution achieved is compared with both the global best and the self-local best of the population. The new position of particles can be obtained by Eqs. (4) and (5).

$$V_i(t+1) = wV_i(t) + c_1 r_1 \big(Pb_i(t) - x_i(t)\big) + c_2 r_2 (G_b(t) - x_i(t)) \quad (4)$$

$$x_i(t+1) = x_i(t) + V_i(t+1), \qquad i = 1,2,\dots,N \quad (5)$$

Where N indicates the number of swarm particles, x_i and V_i represent the position and velocity of the particles respectively, w stands for inertia weight, controlling the influence of the previous velocity on the new one, c_1 and c_2 denote the cognitive and social acceleration coefficient, respectively, and r_1 and r_2 are two random numbers ranging from 0 to 1. It should be noted that, w, c_1, and c_2 can be obtained through performing a trial and error analysis on the dataset under evaluation [88, 89].

2.5. Hybrid machine-learning models developed for Q_g prediction

In this study, four hybrid machine-learning models equipped with effective optimizers are proposed, which provide accurate and reliable predictions of gas flow rate through wellhead chokes. LSSVM and MELM learning algorithms are coupled with two optimization algorithms (GA and PSO) to develop these predictive models.

2.5.1. LSSVM-PSO/GA hybrid models

In this study, two hybrid models LSSVM-PSO and LSSVM-GA, are developed for predicting gas flow rate through the chocks. Fig. 3 displays the flow diagram for the LSSVM- PSO/GA models developed.

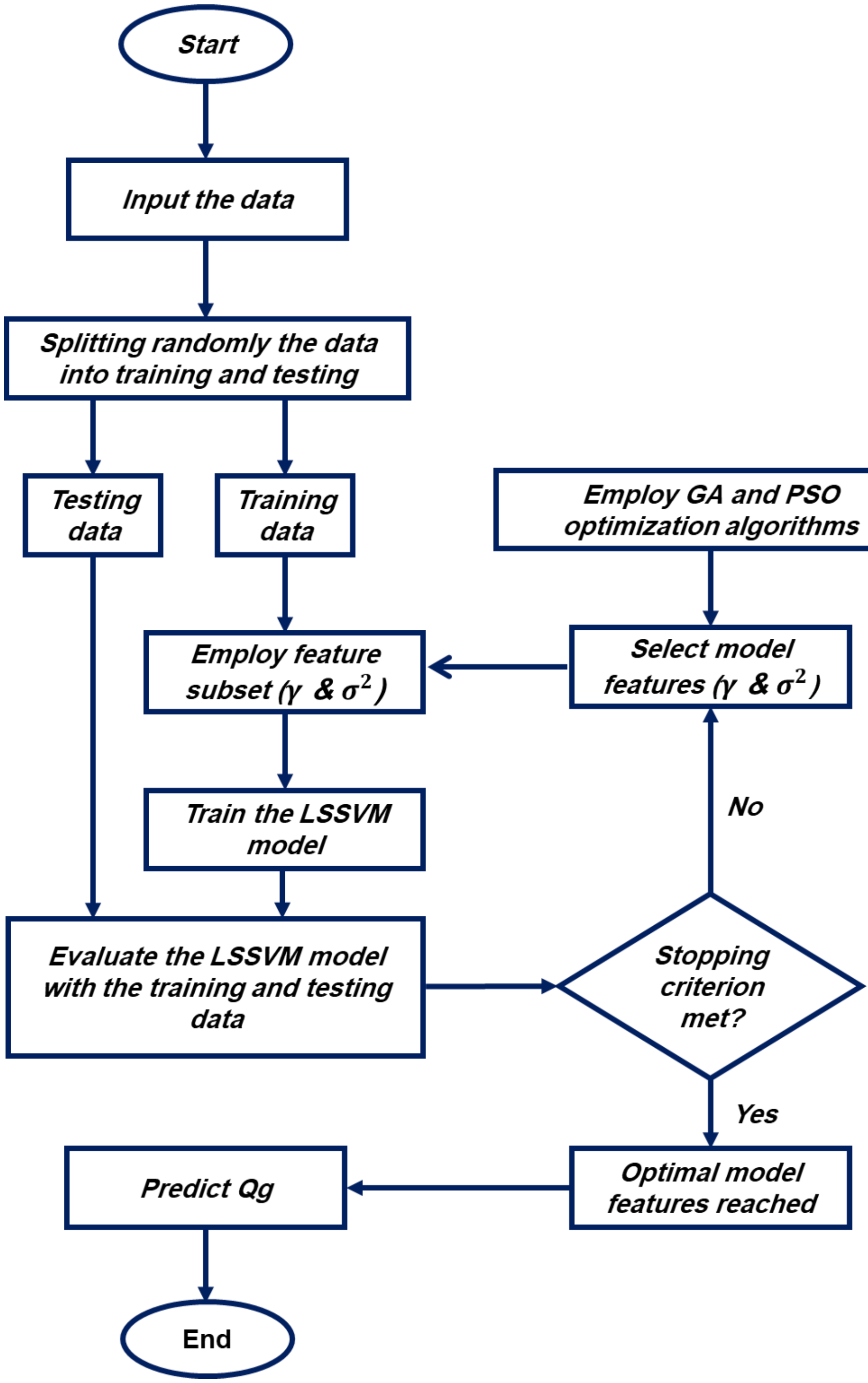

Fig. 3. Typical flow diagram for LSSVR-PSO/GA hybrid models developed for Q_g prediction.

The optimal values of the LSSVM model hyperparameters were obtained using PSO and GA optimization algorithms. RBF kernel function was employed in the LSSVM predictive model construction since it provides the best performance among all the kernel functions tested (table 1). The LSSVM hyperparameters for the hybrid models developed, LLSVM-GA and LSSVM-PSO, and the control parameters for the GA and PSO optimization algorithms applied are listed in Table 3.

Table 3. Optimal values of control parameters for the LSSVM-PSO/GA models established for Q_g prediction.

LSSVM		PSO		GA	
Control parameter	Value	Control parameter	Value	Control parameter	Value
Variance of RBF kernel σ^2	9.8507	Swarm size	80	Population	80
Regularization parameter	53.1392	Maximum iterations	200	Maximum iterations	200
Objective function		Social constant	2.05	Selection method	Roulette wheel
		cognitive constant	2.05	crossover	uniform(p=1)
		Inertia weight	0.98	mutation	uniform(p=1)
				mutation rate	0.08
				selection pressure	2

				(Roulette wheel)	

2.5.2. MELM-PSO/GA hybrid models

Coupling MELM algorithm with GA and PSO optimization, two other hybrid models, MELM-PSO and MELM-GA, were constructed for accurately and reliably predicting gas flow rate through wellhead chokes. The genetic algorithm is inherently discrete, while the PSO algorithm is a continuous method. Both of these algorithms generate new responses in the neighborhood of the two parents (in the genetic algorithm with the crossover operator and the PSO by adsorption to the best position in the P_{best} particle community). Generating answers in the neighborhood of two parents can be one of the most obvious differences with point-based methods such as simulated annealing and taboo search. Execution time in GA is longer than in PSO, and it converges more slowly. The PSO, on the other hand, converges faster due to fewer operators and fewer parameters. More details on the GA and PSO algorithms can be found in previous publications [90-93]. A flow diagram for the MELM-PSO/GA hybrid models developed is illustrated in Fig. 4. As can be seen from Fig. 4, the developed hybrid models include a two-step procedure of optimization, which is briefly described below:

Step1: Determining the optimal number of hidden layers using the optimizers applied by a tuning optimization procedure. The ranges of the numbers of hidden layers and the nodes in those layers are narrowing optimally down. The narrow ranges will then be employed as constraints in constructing hybrid models.

Step 2: Calculating the MELM model's control parameters (weights and biases) for the constrained ranges of the hidden layers and the nodes in those layers obtained at step 1.

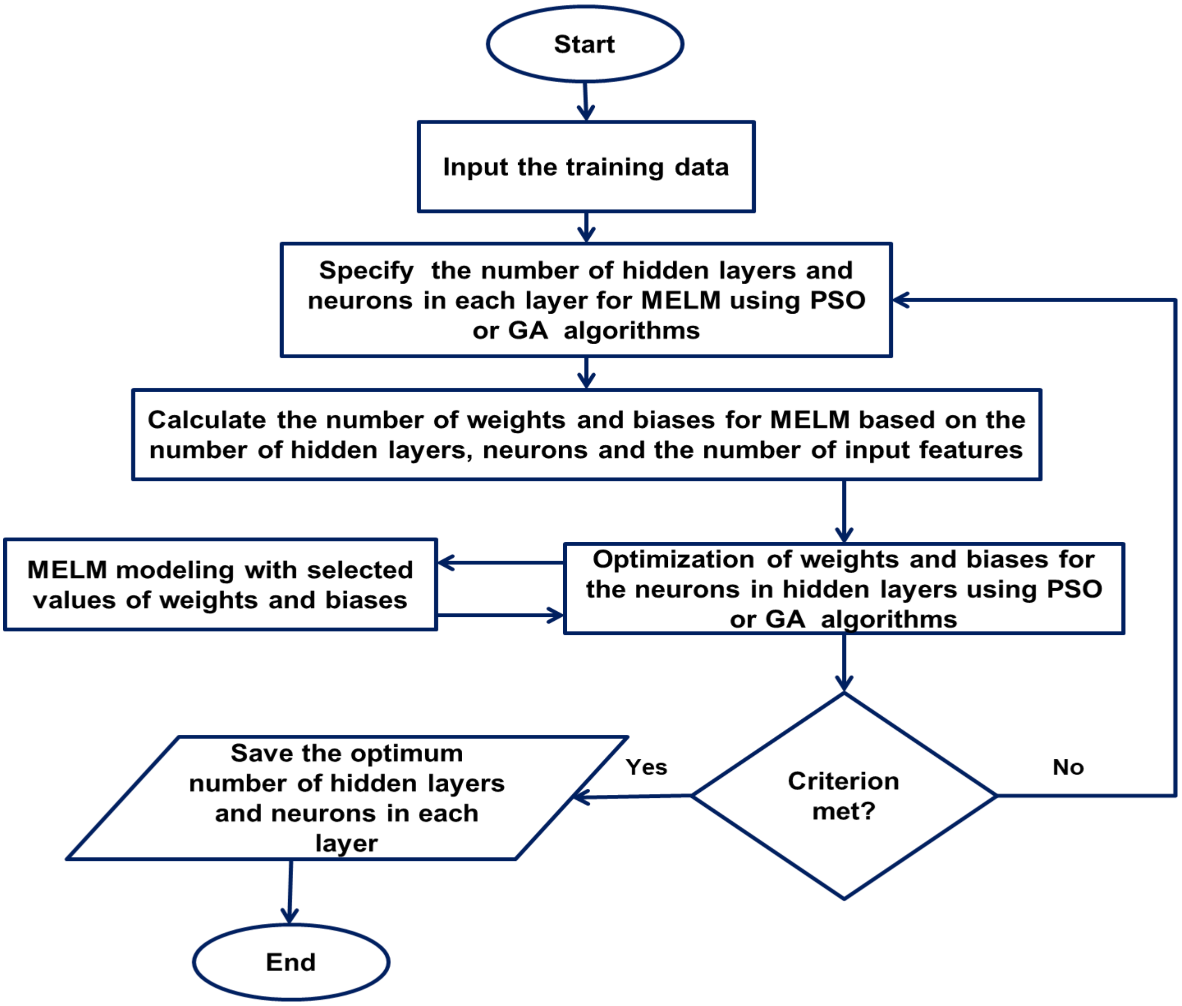

Fig. 4. Typical Flow diagram of MELM-PSO developed for Q_g prediction.

Based on the first step optimization carried out for the MELM construction, the number of hidden layers for MELM is constrained to a range from 5 to 20. The number of nodes in those hidden layers is constrained to a range from 3 to 9. Table 4 lists the results for the first optimization step, and Table 5 shows the best structure for MELM-PSO/GA models. The control parameters for the MELM-PSO/GA hybrid models are presented in Table 6.

Table 4. RMSE obtained for different MELM structures for pre-processing the MELM-PSO/GA models applied for Q_g prediction.

Number of hidden layers	Number of neurons in the layers			
	3	5	7	9
5	6.3296	5.7488	6.0634	6.0985
10	5.8175	5.2953	5.3296	5.3298
15	5.9542	5.0098	5.0108	5.0152
20	5.9533	5.0279	5.0295	5.1841

Table 5. The best structure for pre-processing the MELM-PSO/GA models applied for Q_g prediction.

Layer	Layer1	Layer2	Layer3	Layer4	Layer5	Layer6	RMSE
Neurons	10	9	14	12	12	8	4.9637

Table 6. Optimal values of control parameters for the MELM-PSO/GA hybrid models established for Q_g prediction.

MELM		PSO		GA	
Control parameter	Value	Control parameter	Value	Control parameter	Value
Number of Input variables	6	Swarm size	80	Population	80
Number of hidden layers	20	Maximum iterations	200	Maximum iterations	200
Number of neurons in each layer	5	Social constant	2.05	Selection method	Roulette wheel

	RMSE	cognitive constant	2.05	crossover	uniform(p=1)
		Inertia weight	0.98	mutation	uniform(p=1)
		Var minimum		mutation rate	0.08
		Minimum velocity		selection pressure (Roulette wheel)	2
		Minimum velocity			

3. Data Collection and Distribution

In this paper, for predicting Q_g from gas condensate reservoirs through wellhead chokes, 1067 datasets were collected from three gas condensate fields Marun-Khami, Aghhajari-Khami, and Ahvaz-Khami that located in southwestern Iran (see Fig. 5). Khami group is a group of geological formations of Zagros, which includes Heath and Surmeh formations from the Jurassic period and Fahlian, Gadvan, and Darian formations from the Cretaceous period. This group has crude oil reserves in some oil fields plus gas and condensate gas in most fields. Khami reservoir rock is deeper than the Asmari and Bangestan reservoir rocks. Ahvaz, Gachsaran, Maroon, Karanj, Bibi Hakimeh, and Aghajari fields are among the fields that have crude hydrocarbon reserves (data used in this study are confidential, and the authors have no permission to share them in public). To predict Q_g, six input variables were used in this study, including temperature (T), the upstream pressure (P_u), downstream pressure (P_d), gas gravity (γ_g), choke diameter (D_{64}), and gas-liquid ratio (GLR). To the authors' best knowledge, these six input variables have never been used simultaneously in previously published studies on this topic. Therefore, the models developed in the present study can be considered novel approaches in this field. Table 7 shows the statistical characteristics of the data variables used to predict the Q_g for each reference in this paper.

Table 7. Statistical characterization of data variables in Iranian gas condensate fields for Q_g prediction.

Statistical characterization of the data variables in Iranian gas condensate fields for Qg prediction.								
Field	Variables	Temperature	Upstream Pressure	Downstream Pressure	Gas Specific Gravity	Choke Diameter	Gas Liquid Ratio	Gas Flow Rate
	Symbol	T	Pu	Pd	ɣg	D64	GLR	Qg
	Units	(F)	(Psig)	(Psig)	-	(Inch)	(Scf/STB)	(Mscf/Day)
297 dataset records from Gas Condensate Field (A)	Mean	125.81	1791.70	759.82	0.68	41.94	8.60E+04	20.08
	Std. Deviation	19.67	755.80	350.69	0.04	22.22	7.64E+04	5.99
	Variance	385.67	5.69E+05	1.23E+05	0.00	492.05	5.82E+09	35.76
	Minimum	74.00	217.00	100.00	0.61	16.00	7.46E+03	5.40
	Maximum	187.00	6115.00	2615.00	0.82	160.00	3.22E+05	29.55
	Skewness	0.0894	-0.1461	-0.0105	1.4158	3.0942	1.29E+00	-0.2408
	Kurtosis	-0.1333	3.0393	1.9411	3.1418	10.9981	1.20E+00	-0.9674
	Median	125.00	2043.00	891.00	0.67	40.00	5.76E+04	20.37
	Mode	114.00	2350.00	969.00	0.68	40.00	1.61E+05	28.25
399 dataset records from Gas Condensate Field (B)	Mean	132.18	2045.11	912.01	0.68	128.29	5.98E+04	73.55
	Std. Deviation	22.74	749.42	338.94	0.03	48.86	4.46E+04	17.16
	Variance	515.71	5.60E+05	1.15E+05	0.00	2381.35	1.99E+09	293.70
	Minimum	77.00	1036.00	223.86	0.61	42.00	6.36E+03	54.13
	Maximum	189.00	4658.00	2366.82	0.82	194.00	2.69E+05	122.46
	Skewness	0.0935	1.8326	1.7252	1.3309	-0.1945	1.68E+00	0.8917
	Kurtosis	-0.1346	3.3988	5.0902	3.8104	-1.2896	3.99E+00	-0.2748
	Median	132.00	1880.00	887.00	0.67	130.00	5.36E+04	67.25
	Mode	114.00	1653.00	1003.00	0.69	66.00	8.14E+04	67.41
371 dataset records from Gas Condensate Field (C)	Mean	129.15	2083.33	896.10	0.67	82.99	7.46E+04	42.58
	Std. Deviation	18.91	607.65	322.13	0.04	43.71	6.05E+04	7.16
	Variance	356.56	3.68E+05	1.03E+05	0.00	1905.69	3.66E+09	51.06
	Minimum	85.00	952.00	125.00	0.61	26.00	7.91E+03	29.57
	Maximum	187.00	5910.00	2265.30	0.82	194.00	3.22E+05	53.99
	Skewness	0.1189	2.2510	0.7916	1.4640	1.0655	1.75E+00	-0.1667
	Kurtosis	-0.0658	10.1780	3.6853	3.0654	-0.0534	3.00E+00	-1.1974
	Median	129.00	2088.74	951.00	0.67	64.00	6.18E+04	43.00
	Mode	146.00	1820.00	1070.00	0.68	56.00	8.19E+04	49.44
1067 dataset records from Gas Condensate Fields (A, B and C)	Mean	129.35	1987.86	864.12	0.68	88.50	7.23E+04	47.90
	Std. Deviation	20.76	715.25	342.50	0.04	53.84	6.12E+04	24.68
	Variance	430.67	5.11E+05	1.17E+05	0.00	2896.15	3.75E+09	608.75
	Minimum	74.00	217.00	100.00	0.61	16.00	6.36E+03	5.40
	Maximum	189.00	6115.00	2615.00	0.82	194.00	3.22E+05	122.46
	Skewness	0.1442	1.1531	0.8138	1.3982	0.6632	1.73E+00	0.6162
	Kurtosis	-0.0303	5.0324	3.8315	3.2852	-0.9592	3.21E+00	-0.0940
	Median	129.00	1990.00	908.00	0.67	64.00	5.57E+04	46.84
	Mode	114.00	1986.00	611.50	0.67	64.00	8.14E+04	67.41

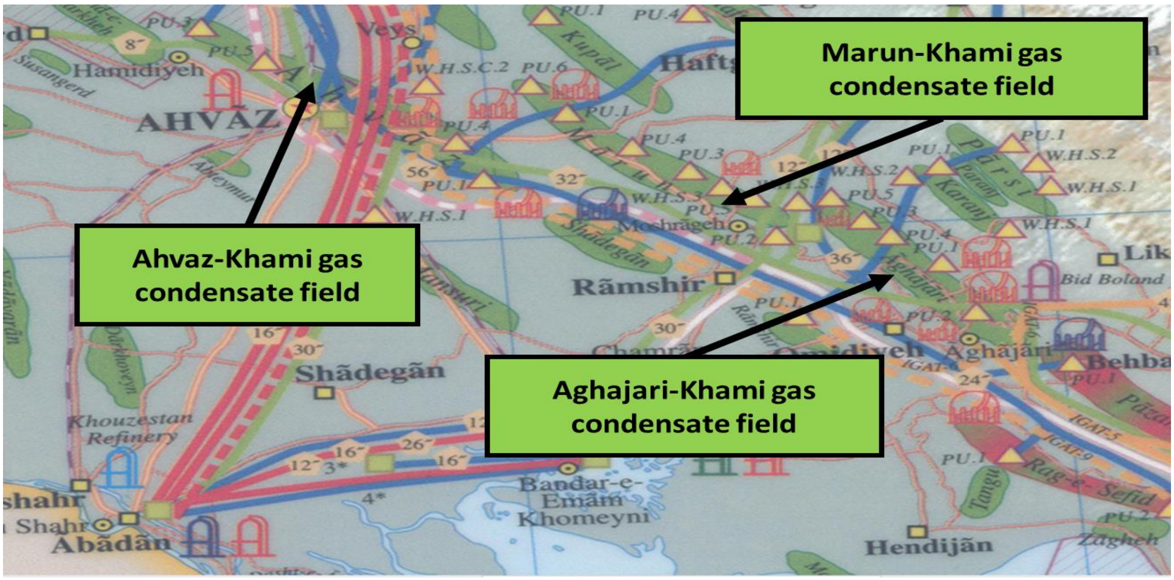

Fig. 5. Marun-Khami, Aghajari-Khami, and Ahvaz-Khami gas condensate fields located onshore Iran in the Zagros Basin.

One of the descriptive diagrams to describe the input data is cumulative distribution functions (CDF) shown in Fig. 6. In this figure (Fig. 6), the 1067 dataset distribution diagram is used, and the CFD formula is shown in Eq. (6) [47, 49]:

$$F_X(x) = P\ (X \leq x), for\ all\ x\epsilon R \qquad (6)$$

X is the data variable value range, *X* is the value of variable x in a specific data record, and *R* is the dataset of data records.

CFD is used to describe the input variables in Fig. 6. The CFD for temperature is T < 112 F^0 for ~ 20.3% of the data records, 112 < T < 152 F^0 for ~ 64.7% of the data records, and T > 152 F^0 for the remaining 15% of the data. The CFD for initial gas specific gravity is γ_g < 0.6588 for ~ 29.8% of the data records, 0.6588 < γ_g < 0.7188 for ~ 64.4% of the data records, and γ_g > 0.7188 for the remaining 5.8% of the data. The CFD for gas to liquid ratio is GLR < 22243 Scf/STB for ~ 21.2% of the data records, 22243 < GLR < 140000 Scf/STB for ~ 60% of the data records, and GLR > 140000 Scf/STB for the remaining 18.8% of the data. The CFD for gas flow rate is Q_g < 18.3

Mscf/Day for ~ 11.3% of the data records, 18.3 < Q_g < 72.8 Mscf/Day for ~ 76.3% of the data records, and Q_g > 72.8 Mscf/Day for the remaining 12.4% of the data. Based on the CFD's shown in Fig. 6, three variable parameters, including T, γ_g, and GLR are normally distributed.

The CFD for upstream pressure is P_u < 2080 psig for ~ 56.4% of the data records, 2080 < P_u < 3220 psig for ~ 39% of the data records, and P_u > 3220 psig remaining 4.6% of the data. The CFD for downstream pressure is P_d < 498.1 psig for ~ 13.67% of the data records, 498.1 < P_d < 966 psig for ~ 48.43% of the data records, 966 < P_d < 1421 psig for ~ 35% of the data records, and P_d > 1421 psig for the remaining 2.9% of the data. The CFD for choke size is D_{64} < 40 inch for ~ 20% of the data records, 40 < D_{64} < 108 inch for ~ 46% of the data records, and D_{64} > 108 inch for the remaining 34% of the data. Based on the CFDs shown in Fig. 6, three variable parameters, including P_u, P_d, and D_{64}, are not normally distributed.

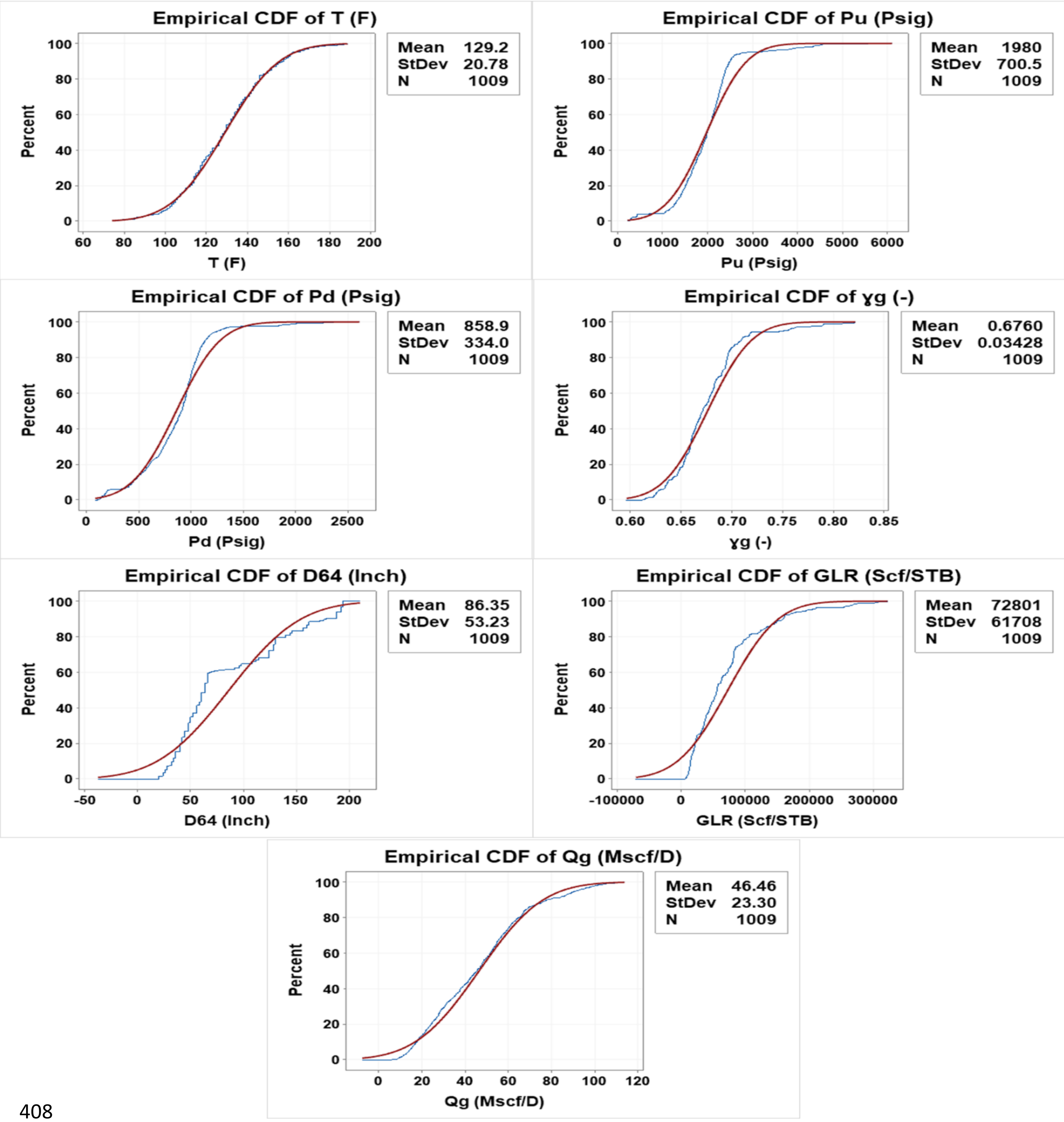

Fig. 6. Cumulative distribution function (CDF) for the input variables and output values used for the Qg prediction (thinner blue line) compared to cumulative

distribution functions for normal distributions defined by variable means and standard deviations (thicker red line).

4. Results & Discussion

Fig. 7 presents the relationship between the input variables (T, Pu, Pd, D64, γ_g, and GLR) and Qg for information on 1009 data records collected around Iran. Comparison of the input variables correlation with Qg indicates that D64 presents a strong correlation with Qg, which suggests this parameter is more influential on Qg than other parameters. Besides, the least influential parameter on the output variable (Qg) is found to be γ_g. This evaluation of the inputs parameters' correlation degree with Qg can assist in the proper selection of features for the algorithms, leading to enhanced prediction performance and accuracy.

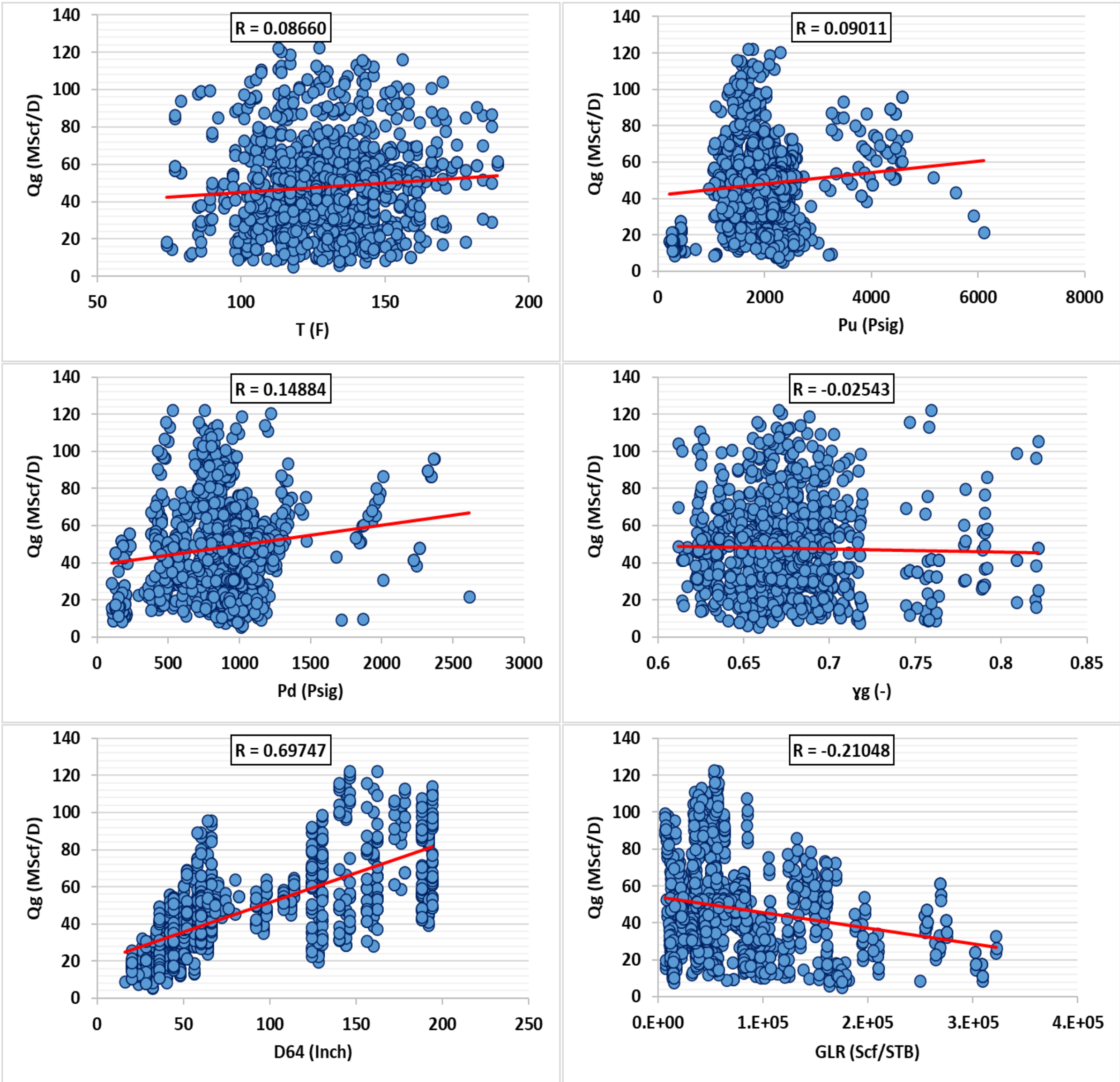

Fig. 7. Cross plot of input variables versus Q_g, indicating the effect of boundaries on the performance of four ML models developed.

One way to compare HML and empirical equations' efficiency in Q_g prediction is to use statistical errors. For this purpose, the equations determining the magnitude of error, including percentage deviation (PD) or relative error (RE), average percentage deviation (APD), absolute average percentage deviation (AAPD), standard deviation (SD), mean square error (MSE), root mean square error (RMSE; the objective function of the HML models), and coefficient of determination (R^2) are selected for prediction accuracy evaluation, which are given in Eqs. (7) to (13):

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

$$PD_i = \frac{H_{(Measured)} - H_{(Predicted)}}{H_{(Measured)}} x\ 100 \qquad (7)$$

Average percentage deviation (APD):

$$APD = \frac{\sum_{i=1}^{n} PD_i}{n} \qquad (8)$$

Absolute average percentage deviation (AAPD):

$$AAPD = \frac{\sum_{i=1}^{n} |PD_i|}{n} \qquad (9)$$

Standard Deviation (SD):

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (D_i - Dimean)^2}{n-1}} \qquad (10)$$

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

Mean Square Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(Z_{Measured_i} - Z_{Predicted_i} \right)^2 \qquad (11)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{MSE} \qquad (12)$$

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (H_{Predicted_i} - H_{Measured_i})^2}{\sum_{i=1}^{N} (H_{Predicted_i} - \frac{\sum_{I=1}^{n} H_{Measured_i}}{n})^2} \qquad (13)$$

These statistical indicators are among the most commonly used indicators to evaluate the prediction performance accuracy and compare HML algorithms and empirical equations. Among these indicators, RMSE is considered the most important one for evaluating HLM models' prediction accuracy. Given these algorithms are configured to minimize the RMSE, this accuracy indicator is more important than other statistical errors studied in this research.

Using statistical errors, the data are divided into two parts: test and train. Tables 8 to 10 show a comparison between the performance accuracy of HML algorithms and empirical models for (712 data records training: 70%), testing (297 data records: 30%), and total subset (1009 data records: 100%) of Iran condensate field data, respectively.

Table 8.* *Gas flow rate Prediction accuracy statistics for the training subset (712 available data records; ~70%) Marun-Khami, Aghajari-Khami, and Ahvaz-Khami gas condensate fields (Q_g; MScf/Day).

Gas flow rate Prediction accuracy statistics for the training subset (712 available data records; ~70%) Marum-Khami, Aghajari-Khami and Ahvaz-Khami gas condensate fields (Qg; MScf/Day).						
Models	**APD**	**AAPD**	**SD**	**MSE**	**RMSE**	**R2**
Units	(%)	(%)	(Mscf/Day)	(Mscf/Day)	(Mscf/Day)	-
Empirical equations						
Osman & Dokla	-93.977	97.968	54.156	5004.2328	70.7406	0.4017
Al-Attar	74.364	83.219	59.382	3991.2958	63.1767	0.4271
Seidi & Sayahi	52.487	61.372	25.974	982.3102	31.3418	0.4952
Ghorbani et al.	31.663	47.116	19.167	602.5881	24.5477	0.4954
Nasriani et al.	47.380	77.087	47.066	2228.3030	47.2049	0.4862
Hybrid machine learning optimizer algorithms						
MELM-PSO	-2.237	5.471	2.592	6.7242	2.5931	0.9900
MELM-GA	-3.179	6.459	3.048	9.3848	3.0635	0.9862
LSSVM-PSO	-5.115	10.870	4.734	22.4867	4.7420	0.9655
LSSVM-GA	-5.194	10.961	5.025	25.3992	5.0398	0.9595

Table 9. ***Gas flow rate Prediction accuracy statistics for the testing subset (297 available data records; ~30%) Marun-Khami, Aghajari-Khami, and Ahvaz-Khami gas condensate fields (Q_g; MScf/Day)***.

Gas flow rate Prediction accuracy statistics for the testing subset (297 available data records; ~30%) Marum-Khami, Aghajari-Khami and Ahvaz-Khami gas condensate fields (Qg; MScf/Day).						
Models	**APD**	**AAPD**	**SD**	**MSE**	**RMSE**	**R2**
Units	(%)	(%)	(Mscf/Day)	(Mscf/Day)	(Mscf/Day)	-
Empirical equations						
Osman & Dokla	-79.112	85.145	47.094	3570.9979	59.7578	0.4392
Al-Attar	69.478	78.021	56.247	3484.8769	59.0328	0.4495
Seidi & Sayahi	55.115	62.059	21.376	924.7682	30.4100	0.4604
Ghorbani et al.	31.160	44.767	18.457	555.9255	23.5781	0.4895
Nasriani et al.	49.449	78.246	46.719	2182.6987	46.7194	0.4651
Hybrid machine learning optimizer algorithms						
MELM-PSO	-3.150	7.220	3.426	11.7437	3.4269	0.9833
MELM-GA	-6.576	12.638	5.840	34.2741	5.8544	0.9508
LSSVM-PSO	-8.134	16.594	6.939	48.2444	6.9458	0.9269
LSSVM-GA	-7.777	15.653	7.051	49.8530	7.0607	0.9241

Table 10. ***Gas flow rate Prediction accuracy statistics for the total subset (1009 available data records; ~100%) Marun-Khami, Aghajari-Khami, and Ahvaz-Khami gas condensate fields (Q_g; MScf/Day)***.

Gas flow rate Prediction accuracy statistics for the total subset (1009 available data records; ~100%) Marum-Khami, Aghajari-Khami and Ahvaz-Khami gas condensate fields (Qg; MScf/Day).						
Models	**APD**	**AAPD**	**SD**	**MSE**	**RMSE**	**R2**
Units	(%)	(%)	(Mscf/Day)	(Mscf/Day)	(Mscf/Day)	-
Empirical equations						
Osman & Dokla	-89.602	94.194	52.391	4582.3589	67.6931	0.4190
Al-Attar	65.448	75.645	53.425	2699.2699	51.9545	0.4239
Seidi & Sayahi	49.657	61.574	24.809	965.3727	31.0704	0.4810
Ghorbani et al.	31.515	46.424	18.964	588.8529	24.2663	0.4905
Nasriani et al.	48.831	77.428	47.004	2214.8793	47.0625	0.4744
Hybrid machine learning optimizer algorithms						
MELM-PSO	-2.506	5.986	2.863	8.2017	2.8639	0.9778
MELM-GA	-4.179	8.278	4.074	16.7110	4.0879	0.9693
LSSVM-PSO	-6.004	12.555	5.476	30.0685	5.4835	0.9534
LSSVM-GA	-5.955	12.342	5.697	32.5972	5.7094	0.9484

Having a close look at the results presented in Tables 8 to 10 reveals that the prediction accuracy of the MELM-PSO algorithm, which is a novel algorithm, is higher than other HML algorithms and empirical equations. For instance, the MELM-PSO model has: RMSE = 2.5931 MScf/Day; AAPD = 5.471٪; R^2 = 0.9900 (for training subset); RMSE = 3.4269 MScf/Day; AAPD = 7.220٪; R^2 = 0.9833 (for testing subset); and RMSE = 2.8639 MScf/Day; AAPD = 5.986٪; R^2 = 0.9778 (for total subset). Besides, HML models are found to be much more efficient than empirical models in terms of prediction accuracy. Comparing the HLM models' prediction performance suggests that comparable prediction accuracy is reached by all four models. Still, the prediction accuracy reached by the MELM-PSO model is slightly higher than those of the MELM-GA and the LSSVM-PSO/GA models.

Fig. 8 shows the Measured versus predicted gas flow rate (Q_g) for each data record in the training, testing, and total subset evaluated for the Iranian condensate fields. Based on the performance accuracy shown in Fig. 8, it is clear that the performance accuracy of HML algorithms is close to each other. In other words, the results of the LSSVM algorithm hybridized with GA / PSO are very close to MELM hybridized with GA / PSO. As shown in Fig. 8, the coefficient of determination value for the MELM-PSO algorithm is much better than other hybrid algorithms. Comparison of the results presented in Tables 8 to 10 and Fig. 8 suggests that the MELM-PSO can achieve higher performance accuracy compared to other models developed in this study. Based on the accuracy, algorithms can be sorted as MELM-PSO > MELM-GA > LSSVM-PSO > LSSVM-GA.

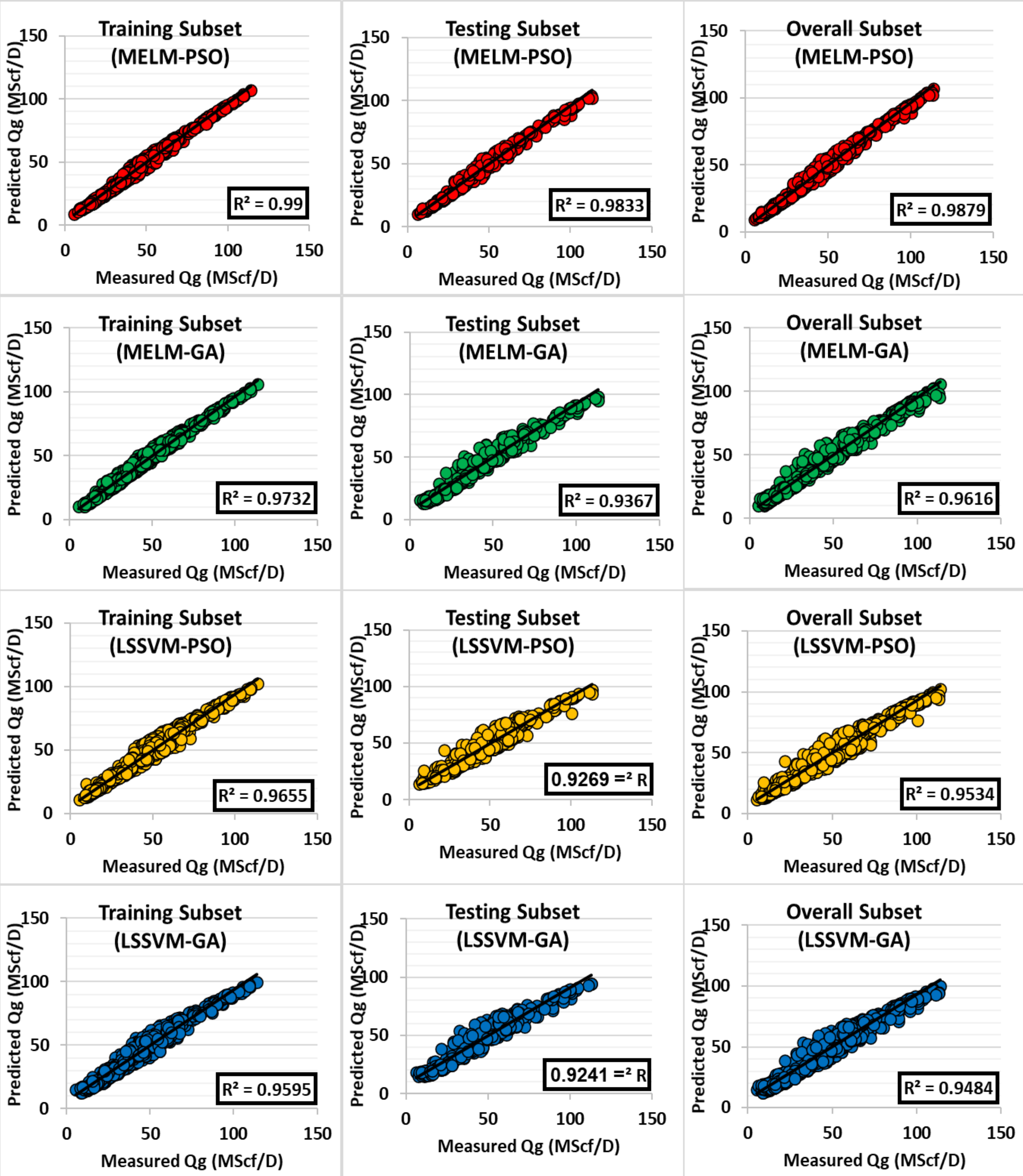

Fig. 8. Measured versus predicted gas flow rate (Q_g) for each data record in the training, testing, and total subset evaluated for HML algorithms (MELM-PSO/GA

and LSSVM-PSO/GA) from the Iranian condensate fields (Marun-Khami, Aghajari-Khami, and Ahvaz-Khami).

Comparison of the results displayed in Figs. 8 and 9 demonstrate that the prediction accuracy of the four HML models developed is much higher than those of previous empirical equations. Based on the prediction accuracy (RMSE), they are as follows: MELM-PSO > MELM-GA > LSSVM-PSO > LSSVM-GA > Ghorbani et al. > Seeidi & Sayahi > Nasriani et al. > Al-Attar > Osman & Dokla.

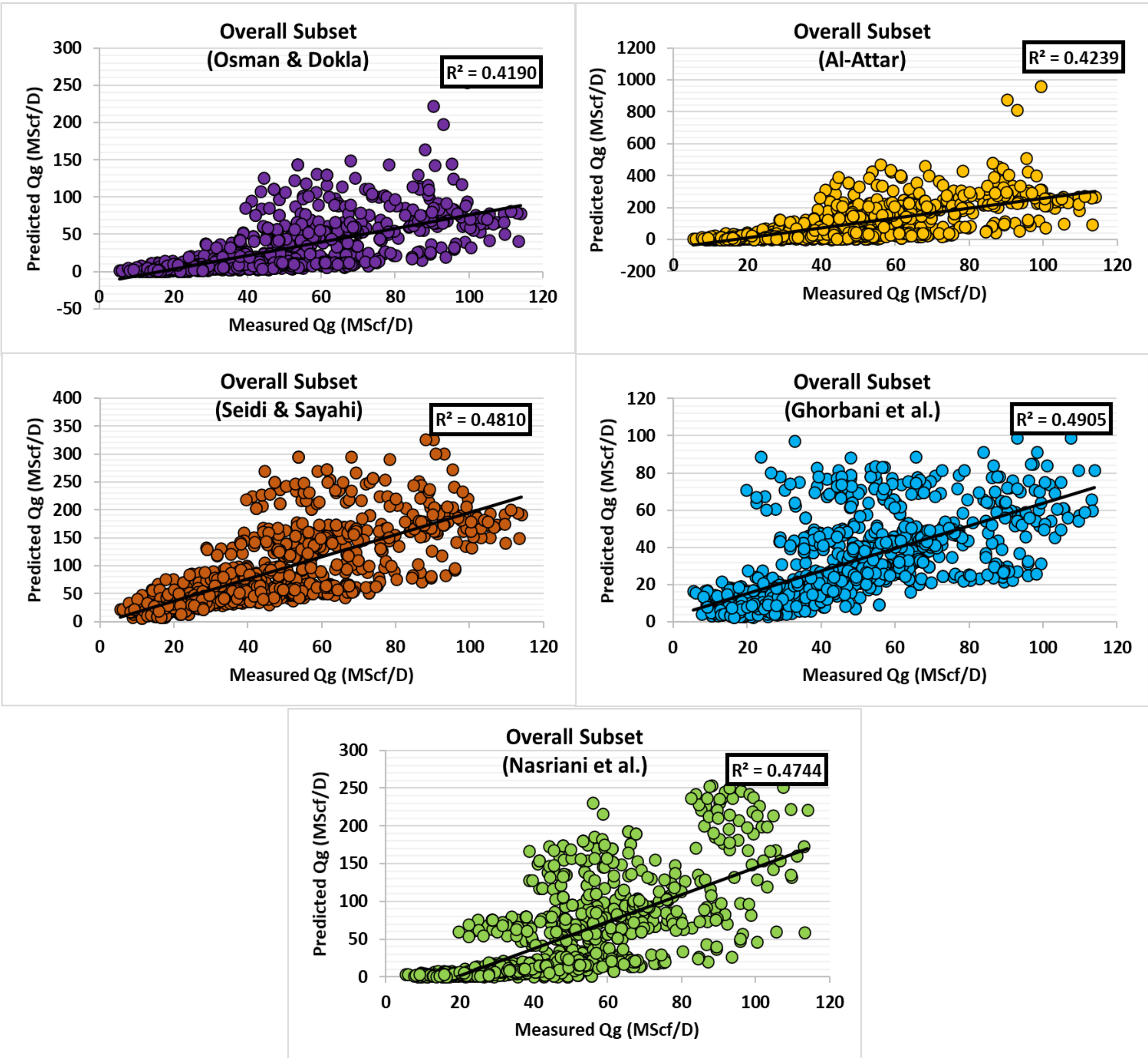

Fig. 9. Measured versus predicted gas flow rate (Q_g) for each data record in the training, testing, and total subset evaluated for empirical equations (Osman & Dokla, Al-Attar, Seeidi & Sayahi, Ghorbani et al., and Nasriani et al.) from the Iranian condensate fields (Marun-Khami, Aghajari-Khami, and Ahvaz-Khami).

Figs. 10 and 11 display the histograms of gas flow rate prediction error with normal distributions (red line) for the HML algorithms and the empirical equations based on 1009 subset data records from the Iranian condensate fields. As shown in Fig. 10, the error rate for the HML is close to zero, and the lowest error for these models is obtained by MELM-PSO. However, the error for all empirical equations is shifted to the right (Fig. 11). According to the results of this figure (Fig. 11), it is clear that the error distribution for the experimental models Osman & Dokla and Al-Attar is asymmetric. All the empirical models involve some individual predictions involving quite large errors, particularly in the positive direction (i.e., overestimates of Q_g). The lowest Q_g prediction error range is associated with is MELM-PSO model.

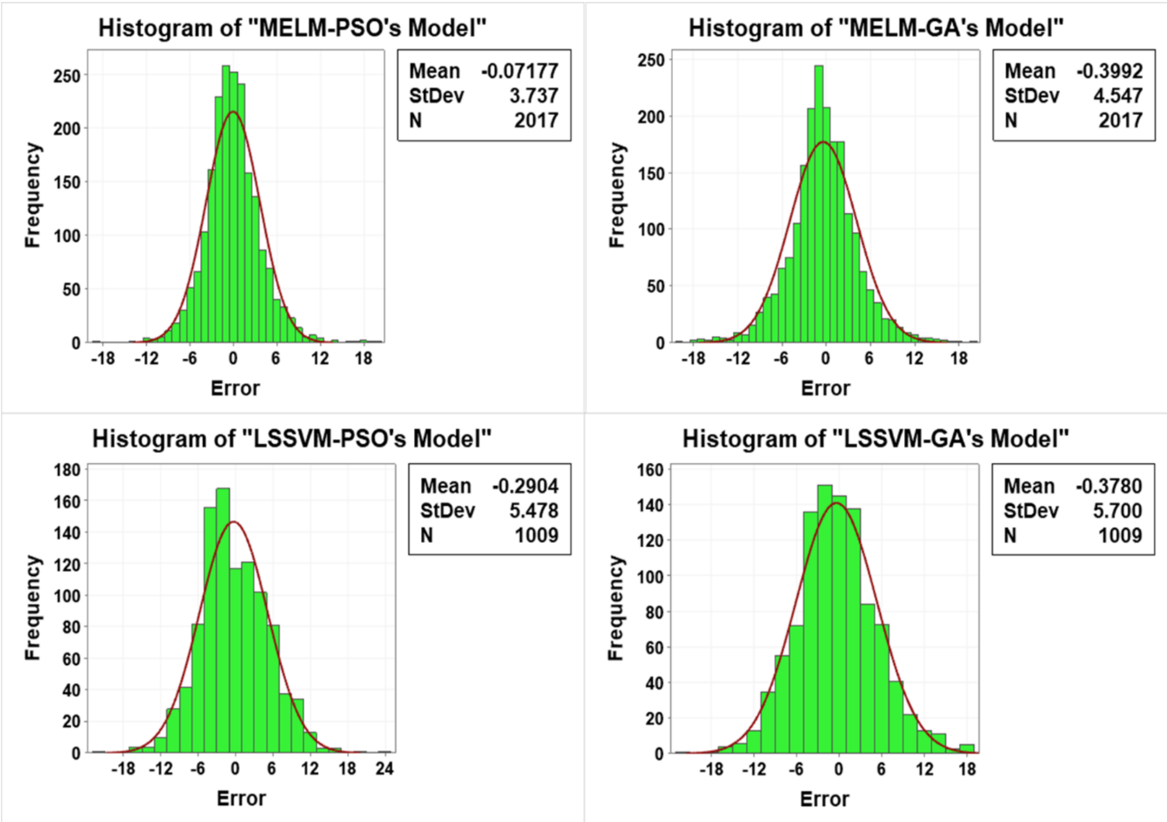

Fig. 10. Gas flow rate prediction error (Q_g) histograms displayed with normal distributions (red line) for HML algorithms based on 1009 subset data records from the Iranian condensate fields (Marun-Khami, Aghajari-Khami, and Ahvaz-Khami).

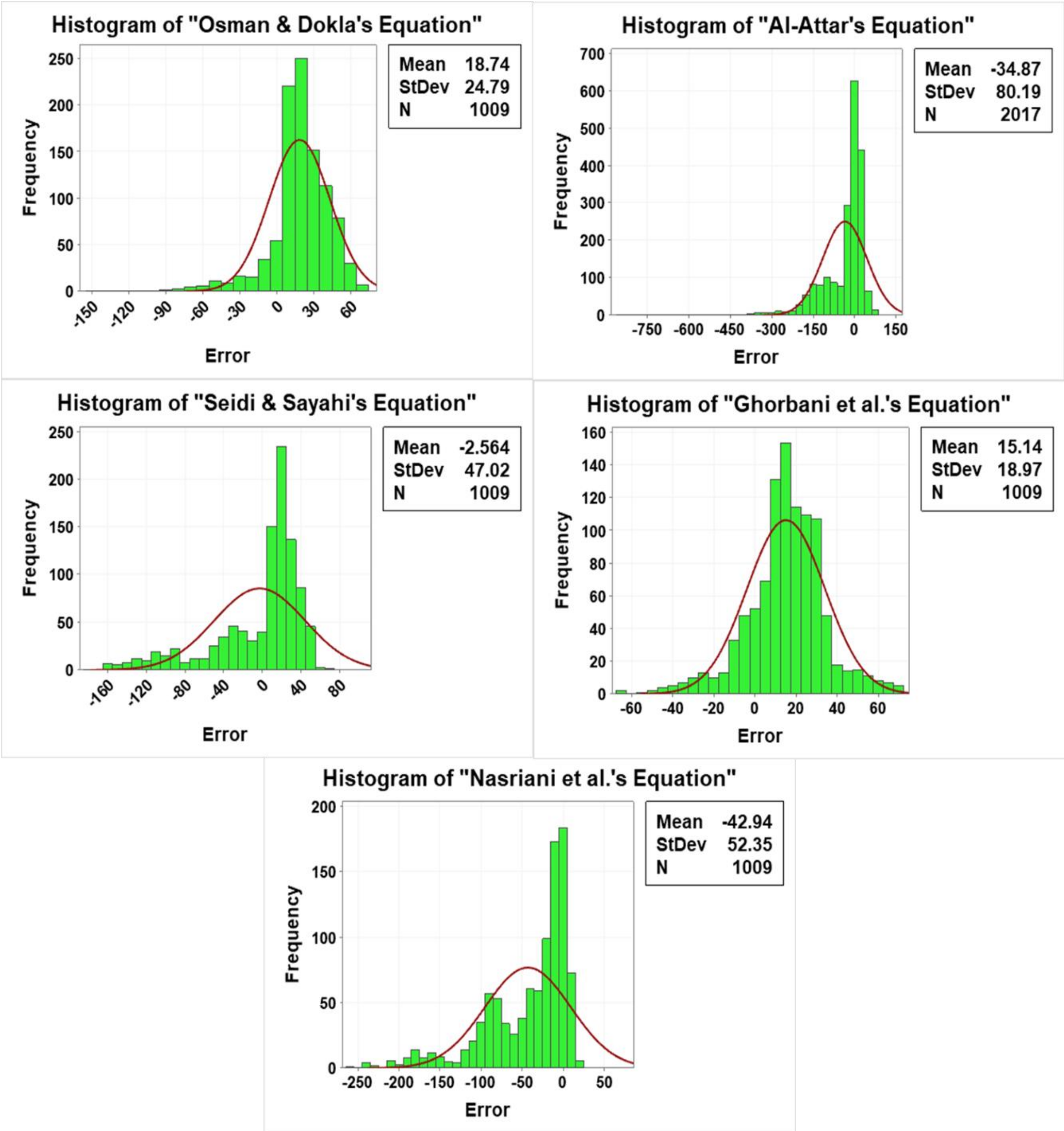

Fig. 11. Gas flow rate prediction error (Q_g) histograms displayed with normal distributions (red line) for empirical equations based on 1009 subset data records from the Iranian condensate fields (Marun-Khami, Aghajari-Khami, and Ahvaz-Khami).

One of the most important and influential factors on the performance accuracy of a prediction model is the use of high-quality data [94, 95]. However, due to the lack of calibration of measuring devices, field data always presents a degree of errors [38]. In other words, there can be data recodes among datasets that are far from the truth. These poor-quality data cause problems in the machine learning process and the training model built on artificial intelligence.

When dealing with such data, identifying and deleting unreliable data with distinct outlying values is the best way to increase the model's accuracy. To identify and remove erroneous data parenting in the dataset under study, K-means clustering method in a multidimensional space is used. For this purpose, two to five clusters are considered, which are then divided into smaller clusters [96]. Remote data sets are used as part of the data processing phase to input data into a single-layer ANN network with five neurons to predict Q_g.

The results of the K-means clustering performed are shown in Fig. 12. As it can be seen, 3 clusters demonstrate the lowest RMSE for Q_g prediction. Based on this modeling, 58 data sets are identified as outlier data sets. The K-means clustering algorithm can retrieve remote data to predict Qg. Fig. 13 displays that the k-means clustering presents a promising efficiency in outlier detection for the prediction of Q_g.

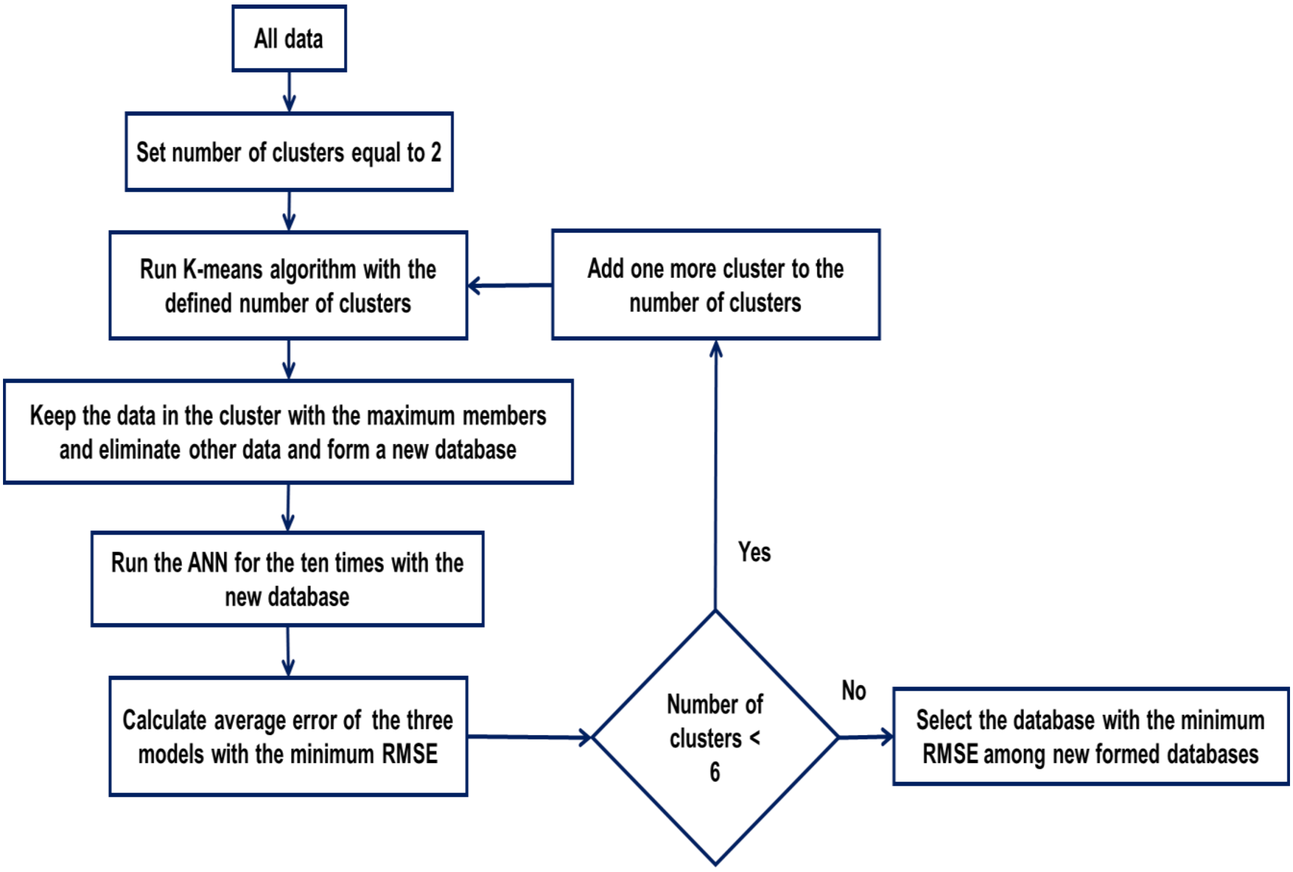

Fig. 12. Schematic of identifying and deleting to data outlier detection using the K-means clustering algorithm [38].

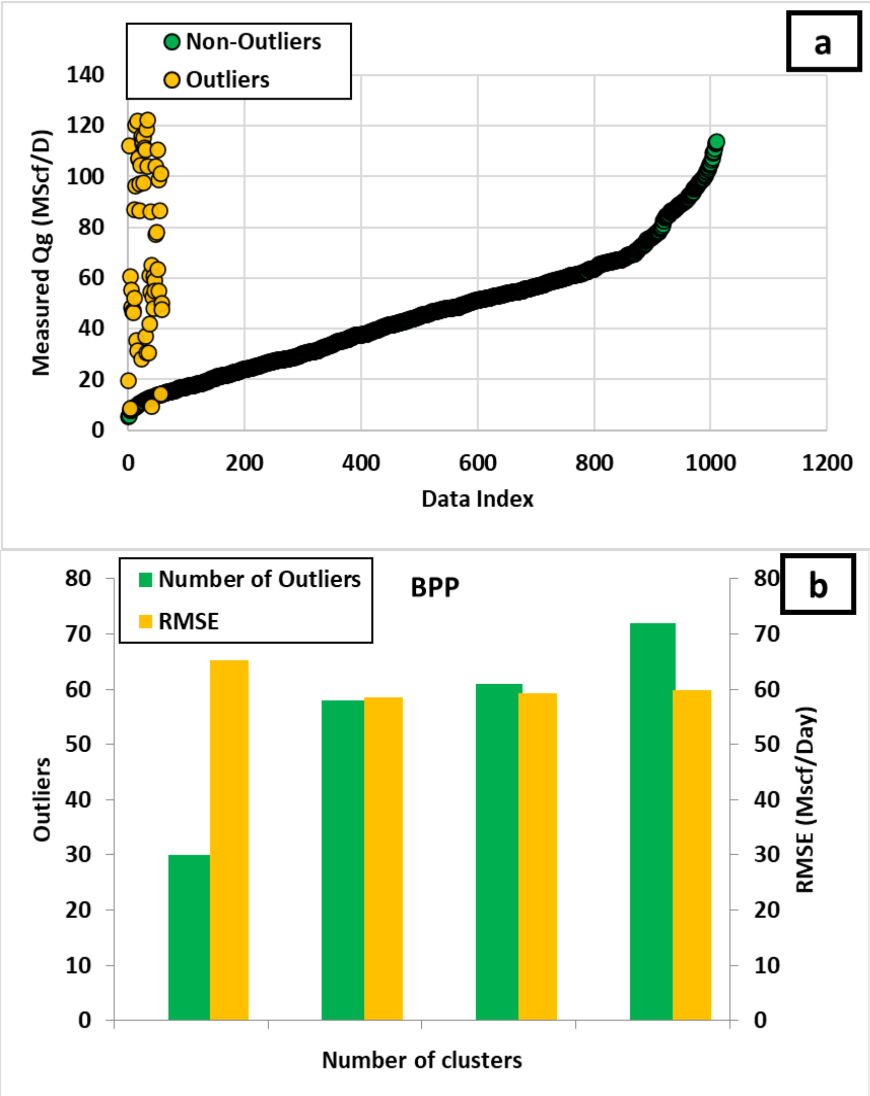

Fig. 13. Results of outlier detection by the K-means clustering algorithm a) Status of remote data detected for Q_g prediction and b) Number of outlying data detected per number of different clusters and ANN modeling error after removal of remote data to predict Q_g.

Fig. 14 demonstrates how the HML models developed progress towards optimal and accurate prediction of Qg through two hundred iterations. Comparing the results displayed in Fig. 14 indicates that all four HLM algorithms present relatively similar convergence velocity in iteration #3. As seen in iteration #86, the prediction accuracy of LSSVM-PSO is better than that of LSSVM-GA. As for MELM-PSO/GA models, PSO presents a quicker convergence to achieve its best solution than the GA optimizer. From iteration #120, the MELM-PSO performs better than the MELM-GA in terms of prediction accuracy. All in all, the MELM-PSO/GA models are found to present higher forecast accuracy than those of the LSSVM-PSO/GA. In addition, the PSO optimizer is found to be more efficient in reaching the optimal solution for both networks, the MELM and the LSSVM, when compared to the GA optimizer.

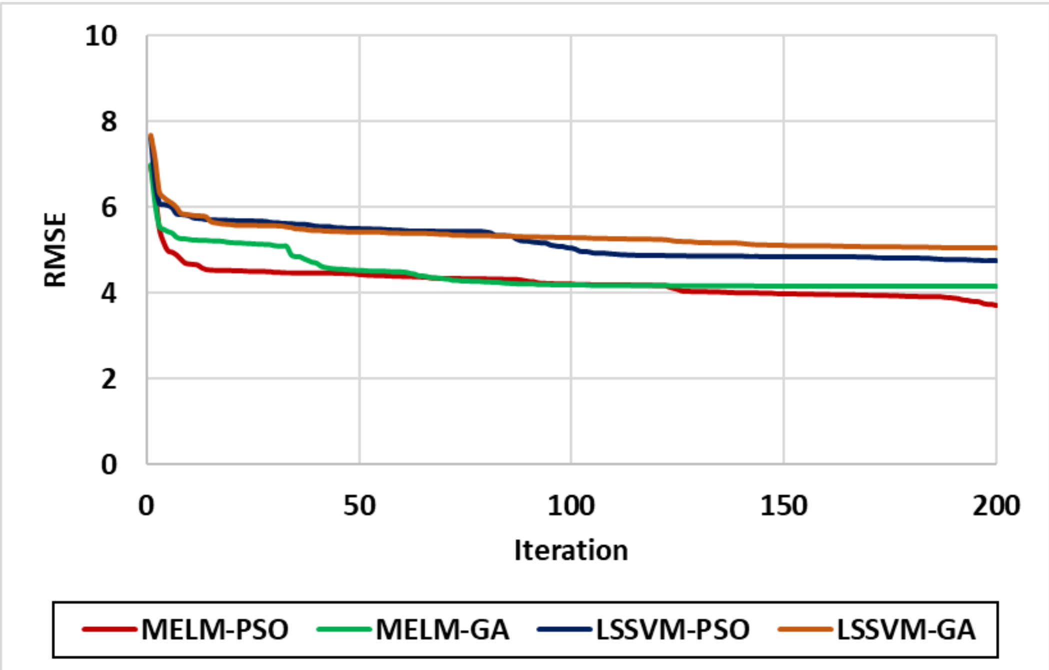

Fig. 14. RMSE values for the training subset based on HML algorithms (MELM-PSO, MELM-GA, LSSVM-PSO, and LSSVM-GA) developed for the prediction of Q_g during supervised learning from the Iranian condensate fields (Marun-Khami, Aghajari-Khami, and Ahvaz-Khami).

To determine the degree of influence of each input variable on Qg, Spearman's non-parametric correlation coefficient (ρ) is used [97]. The range of this parameter is between −1 (complete negative correlation) to 1 (complete positive correlation), which indicates a relatively low or high impact [98]. The Spearman parameter equation (Eq. (14)) is defined as follows:

$$\rho = \frac{\sum_{i=1}^{n}(E_i - \bar{E})(F_i - \bar{F})}{\sqrt{\sum_{i=1}^{n}(E_i - \bar{E})^2 \sum_{i=1}^{n}(F_i - \bar{F})^2}} \tag{14}$$

Where E_i is E input variable value of data record I, $\bar{E}$ is mean value for variable E, F_i is F dependent variable (Q_g) value of data record I, $\bar{F}$ is mean value for dependent variable F, and n is the number of input parameters.

Fig. 15 shows the calculated ρ value for the total of 1009 processed learning datasets. Based on the correlation coefficients determined, it is observed that D_{64}, P_d, P_u, and T parameters positively influence Q_g, whereas GLR and γg parameters present a negative influence on it. The greatest positive influence on Qg is observed for D_{64}, while the greatest negative influence is presented by GLR (see in Eq. (15)). In general, the order of input variables' influence degree on Qg is as follows: choke diameter (D_{64}) > downstream pressure (P_d) > gas-liquid ratio (GLR) > upstream pressure (P_u) > temperature (T) > gas gravity (γ_g).

$$Q_g \propto (D_{64}, P_d, P_u, T) \quad \text{and} \quad Q_g \propto \frac{1}{(GLR, \gamma_g)} \tag{15}$$

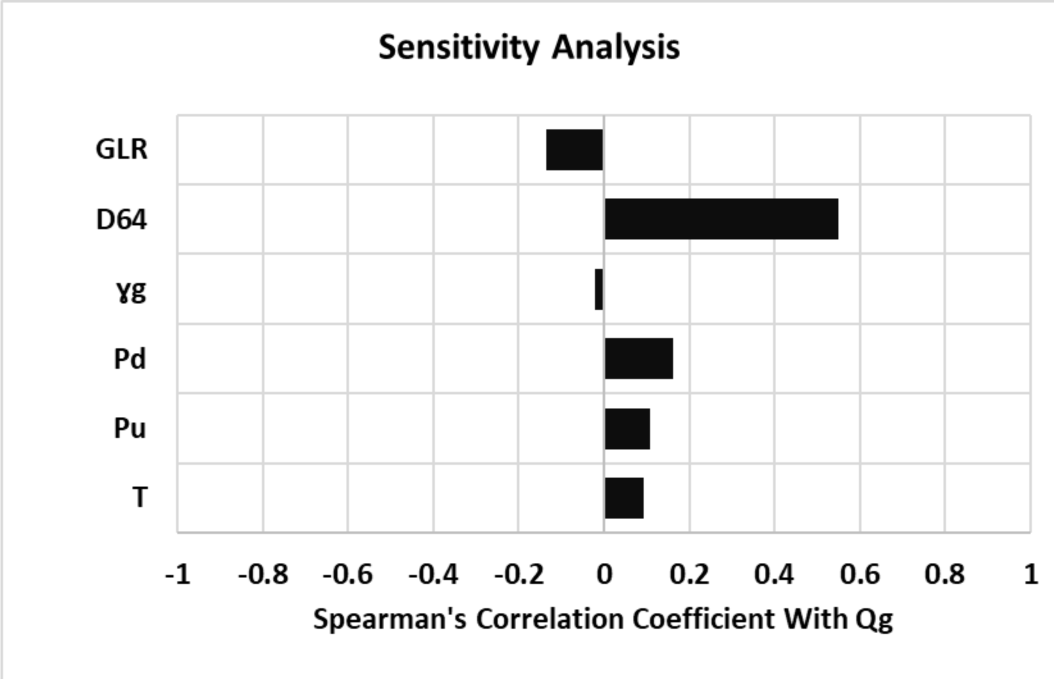

Fig. 15. Input variables assessed based on Spearman's non-parametric correlation coefficient values for Q_g prediction calculated for 1009 data records of supervised learning dataset (from Iranians condensate fields (Marun-Khami, Aghajari-Khami, and Ahvaz-Khami)).

5. Conclusion

In this research, 1009 input data from Iranian condensate fields (Marun-Khami, Aghajari-Khami, and Ahvaz-Khami) are used to construct four models to predict gas flow rate (Q_g) through six input variables. The input variables to the developed models are temperature (T), the upstream pressure (P_u), downstream pressure (P_d), gas gravity (γ_g), choke diameter (D_{64}), and gas-liquid ratio (GLR). This is the first-ever research work constructing a model based on these variables.

Hybrid machine learning algorithms have several advantages over simple machine learning algorithms. For instance, when the predictive machine learning algorithms are combined with the PSO algorithm to determine control parameters of the algorithms, the computational speed and accuracy enhance remarkably. In the case of the MELM model, they are optimized in two steps. The first step is to determine the number of hidden layers and neurons in the network. The next is to identify the desired weight and biases applied to those layers and neurons. In the case of LSSVM, the optimization setting is done in one step for the development of LSSVM with PSO/GA optimizer, which ultimately leads to LSSVM-PSO and LSSVM-GA hybrid machine learning optimizer algorithms.

Coupling the PSO to the GA algorithm is an effective approach in achieving high prediction accuracy in the HML algorithms. The multi-hidden layer extreme learning machine (MELM) algorithm coupled with the PSO optimizer presents the best performance. This algorithm uses two hybrid stages with PSO to improve its performance. This algorithm (MELM) first reduces the number of layers and nodes in each hidden layer by combining with PSO. In combination with the second PSO, determines the appropriate weight and bias for the nodes of the selected hidden layers.

The performance accuracy obtained by the MELM-PSO model applied to the total subset entered is RMSE = 2.8639 MScf/Day and R^2 = 0.9778, which is significantly higher than the prediction accuracy of empirical equations and HML models. The best performance accuracy obtained from Empirical equations related to Ghorbani et al., Which is RMSE = 24.2663 MScf / Day and R^2 = 0.4905. Comparing the developed MELM-PSO model with the previous empirical (Table 1), the AI models (Table 2) suggest that the MELM-PSO model has superior prediction performance and higher accuracy.

Sensitivity analysis obtained from the Spearman coefficient model demonstrates that the input variables, including D_{64}, P_d, P_u, and T, have positive correlations with Q_g. In contrast, GLR and γ_g parameters present negative correlations with Q_g. D_{64} displays the greatest positive correlation with Q_g, whereas the poorest negative correlation with Q_g is observed for GLR.

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.

Acknowledgment

This research was supported by Tomsk Polytechnic University development program.

Nomenclature

ANN	=	Artificial Neural Network
ANFIS	=	Adaptive Neuro-Fuzzy Inference System
b	=	Bias vector
BP	=	Backpropagation
CF	=	Cost Function
CFD	=	Cumulative distribution functions
c1	=	Positive cognitive coefficient (individual learning factors PSO)
c2	=	Positive social coefficient (global learning factor for PSO)
d	=	The degree of polynomial
D64	=	Choke size
DL	=	Deep learning
$\bar{E}$	=	Mean value for variable E
Ei	=	Input variable value of data record i
ELM	=	Extreme Learning Machine
$\bar{F}$	=	Mean value for dependent variable F
Fi	=	Input variable value of data record i
FN	=	Functional Network
GA	=	Genetic algorithm
G_b	=	The global best value found in the swarm

GEP	=	Gene expression programming
GLR	=	Gas to liquid ratio
LSSVM	=	Least Squares Support Vector Machine
M, l, O	=	Experimental coefficients
MELM	=	Multiple Extreme Learning Machine
MLP	=	Multi-Layer Perceptron
N	=	Number of samples in dataset
n	=	Number of inputs parameters
PSO	=	Particle swarm optimization
P_b	=	The cognitive best value of particle
P_{wh}	=	Wellhead pressure
Pd	=	Downstream pressure
Pu	=	Upstream pressure
Q_g	=	Gas flow rate
Q_{liq}	=	Rate of liquids production
RBF	=	Radial basis function
RMSE	=	Root mean square error
SVM	=	Support Vector Machines
T	=	Transpose matrix
t	=	The intercept of polynomial
y_i	=	Output vector
V_i	=	Particle ith velocity in PSO swarm
W	=	Inertial weight (PSO)
w	=	Weight vector
a_i	=	Lagrangian function multiplier
e_i	=	Regression error
X_i	=	Particle i^{th} position in PSO swarm
x_i	=	Input variable
σ^2	=	The variance of Gaussian kernel
γ	=	Adjustable factor
Δp	=	Differential pressure
θ and k	=	Bias and scale parameters
$\phi(x_i)$	=	Kernel function

Refrences

[1] Amirian E, Dejam M, Chen Z. Performance forecasting for polymer flooding in heavy oil reservoirs. Fuel 2018;216:83-100. doi: https://doi.org/10.1016/j.fuel.2017.11.110.

[2] Litvinenko V. The role of hydrocarbons in the global energy agenda: The focus on liquefied natural gas. Resources 2020;9(5):59. doi: https://doi.org/10.3390/resources9050059.

[3] Zou C, Zhao Q, Zhang G, Xiong B. Energy revolution: From a fossil energy era to a new energy era. Natural Gas Industry B 2016;3(1):1-11. doi: https://doi.org/10.1016/j.ngib.2016.02.001.

[4] Song X, Liu Y, Xue L, Wang J, Zhang J, Wang J, et al. Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. Journal of Petroleum Science and Engineering 2020;186:106682. doi: https://doi.org/10.1016/j.petrol.2019.106682.

[5] Yadua AU, Lawal KA, Okoh OM, Ovuru MI, Eyitayo SI, Matemilola S, et al. Stability and stable production limit of an oil production well. Journal of Petroleum Exploration and Production Technology 2020;10(8):3673-87. doi: https://link.springer.com/article/10.1007/s13202-020-00985-3.

[6] Naik G. Tight gas reservoirs–an unconventional natural energy source for the future. Accessado em 2003;1(07):2008. doi: www.pinedaleonline.com/socioeconomic/pdfs/tight_gas.pdf.

[7] El-Banbi AH, McCain Jr W, Semmelbeck M. Investigation of well productivity in gas-condensate reservoirs. *SPE/CERI Gas Technology Symposium.* Society of Petroleum Engineers; 2000. doi: https://doi.org/10.2118/59773-MS.

[8] Hekmatzadeh M, Gerami S. A new fast approach for well production prediction in gas-condensate reservoirs. Journal of Petroleum Science and Engineering 2018;160:47-59. doi: https://doi.org/10.1016/j.petrol.2017.10.032.

[9] Alavi F, Mowla D, Esmaeilzadeh F. Production performance analysis of Sarkhoon gas condensate reservoir. Journal of Petroleum Science and Engineering 2010;75(1-2):44-53. doi: https://doi.org/10.1016/j.petrol.2010.10.002.

[10] Mokhtari R, Varzandeh F, Rahimpour M. Well productivity in an Iranian gas-condensate reservoir: a case study. Journal of Natural Gas Science and Engineering 2013;14:66-76. doi: https://doi.org/10.1016/j.jngse.2013.05.006.

[11] Janiga D, Czarnota R, Stopa J, Wojnarowski P, Kosowski P. Utilization of nature-inspired algorithms for gas condensate reservoir optimization. Soft Computing 2019;23(14):5619-31. doi: https://link.springer.com/article/10.1007/s00500-018-3218-6.

[12] Hassan AM, Mahmoud MA, Al-Majed AA, Al-Shehri D, Al-Nakhli AR, Bataweel MA. Gas Production from Gas Condensate Reservoirs Using Sustainable Environmentally Friendly Chemicals. Sustainability 2019;11(10):2838. doi: https://doi.org/10.3390/su11102838.

[13] Kaydani H, Najafzadeh M, Mohebbi A, Practice. Wellhead choke performance in oil well pipeline systems based on genetic programming. Journal of Pipeline Systems Engineering 2014;5(3):06014001. doi: https://doi.org/10.1061/(ASCE)PS.1949-1204.0000165.

[14] Hansen LS, Pedersen S, Durdevic P. Multi-phase flow metering in offshore oil and gas transportation pipelines: Trends and perspectives. Sensors 2019;19(9):2184. doi: https://doi.org/10.3390/s19092184.

[15] Mokhtari K, Waltrich P. Performance evaluation of multiphase flow models applied to virtual flow metering. WIT Transactions on Engineering Sciences 2016;105:99-111. doi:
[16] Ki S, Jang I, Cha B, Seo J, Kwon O. Restoration of Missing Pressures in a Gas Well Using Recurrent Neural Networks with Long Short-Term Memory Cells. Energies 2020;13(18):4696. doi: https://doi.org/10.3390/en13184696.
[17] Lak A, Azin R, Osfouri S, Gerami S, Chahshoori R. Choke modeling and flow splitting in a gas-condensate offshore platform. Journal of Natural Gas Science and Engineering 2014;21:1163-70. doi: https://doi.org/10.1016/j.jngse.2014.07.020.
[18] Ghorbani H, Moghadasi J, Wood DA. Prediction of gas flow rates from gas condensate reservoirs through wellhead chokes using a firefly optimization algorithm. Journal of Natural Gas Science and Engineering 2017;45:256-71. doi: https://doi.org/10.1016/j.jngse.2017.04.034.
[19] Chong D, Yan J, Wu G, Liu J. Structural optimization and experimental investigation of supersonic ejectors for boosting low pressure natural gas. Applied Thermal Engineering 2009;29(14-15):2799-807. doi: https://doi.org/10.1016/j.applthermaleng.2009.01.014.
[20] Schüller R, Solbakken T, Selmer-Olsen S, facilities. Evaluation of multiphase flow rate models for chokes under subcritical oil/gas/water flow conditions. SPE production 2003;18(03):170-81. doi: https://doi.org/10.2118/84961-PA.
[21] Coutinho RP, Waltrich PJ, Williams WC, Mehdizadeh P, Scott S, Xu J, et al. Experimental Characterization of Two-Phase Flow Through Valves Applied to Liquid-Assisted Gas-Lift. Journal of Energy Resources Technology 2020;142(6). doi: https://doi.org/10.1115/1.4045921.
[22] Gilbert W. Flowing and gas-lift well performance. *Drilling and production practice.* American Petroleum Institute; 1954. doi: https://onepetro.org/APIDPP/proceedings-abstract/API54/All-API54/API-54-126/51072.
[23] Safar Beiranvand M, Mohammadmoradi P, Aminshahidy B, Fazelabdolabadi B, Aghahoseini S. New multiphase choke correlations for a high flow rate Iranian oil field. Mechanical Sciences 2012;3(1):43-7. doi: https://doi.org/10.5194/ms-3-43-2012.
[24] Aladwani F, Alatefi S. Toward the Development of a Universal Choke Correlation–Global Optimization and Rigorous Computational Techniques. Journal of Engineering Research 2020;8(3). doi: https://doi.org/10.36909/jer.v8i3.7717
[25] Bennis M, Gellert J, Nougués M, Crespo P. Decline Curve Analysis in Vaca Muerta with Choke Size Normalization of Gas Rates. *SPE/AAPG/SEG Latin America Unconventional Resources Technology Conference.* Unconventional Resources Technology Conference; 2020. doi: https://doi.org/10.15530/urtec-2020-1403.
[26] Khamis M, Elhaj M, Abdulraheem A. Optimization of choke size for two-phase flow using artificial intelligence. Journal of Petroleum Exploration and Production Technology 2020;10(2):487-500. doi: https://link.springer.com/article/10.1007/s13202-019-0734-6.
[27] Osman ME, Dokla ME. Gas condensate flow through chokes. *European Petroleum Conference.* Society of Petroleum Engineers; 1990. doi: https://doi.org/10.2118/20988-MS.

[28] Guo B, Al-Bemani AS, Ghalambor A. Applicability of Sachdeva's choke flow model in Southwest Louisiana gas condensate wells. *SPE Gas Technology Symposium.* Society of Petroleum Engineers; 2002. doi: https://doi.org/10.2118/75507-MS.
[29] Al-Attar H. Performance of wellhead chokes during sub-critical flow of gas condensates. Journal of Petroleum Science and Engineering 2008;60(3-4):205-12. doi: https://doi.org/10.1016/j.petrol.2007.08.001.
[30] Nasriani HR, Khan K, Graham T, Ndlovu S, Nasriani M, Mai J, et al. An Investigation into Sub-Critical Choke Flow Performance in High Rate Gas Condensate Wells. Energies 2019;12(20):3992. doi: https://doi.org/10.3390/en12203992.
[31] Seidi S, Sayahi T. A new correlation for prediction of sub-critical two-phase flow pressure drop through large-sized wellhead chokes. Journal of Natural Gas Science and Engineering 2015;26:264-78. doi: https://doi.org/10.1016/j.jngse.2015.06.025.
[32] Hassanpouryouzband A, Joonaki E, Edlmann K, Haszeldine RS. Offshore Geological Storage of Hydrogen: Is This Our Best Option to Achieve Net-Zero? ACS Energy Letters 2021:2181-6. doi: https://doi.org/10.1021/acsenergylett.1c00845.
[33] Hassanpouryouzband A, Yang J, Okwananke A, Burgass R, Tohidi B, Chuvilin E, et al. An Experimental Investigation on the Kinetics of Integrated Methane Recovery and CO 2 Sequestration by Injection of Flue Gas into Permafrost Methane Hydrate Reservoirs. Scientific reports 2019;9(1):1-9. doi: https://doi.org/10.1038/s41598-019-52745-x.
[34] Hassanpouryouzband A, Yang J, Tohidi B, Chuvilin E, Istomin V, Bukhanov B. 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 Sustainable Chemistry and Engineering 2019;7(5):5338-45. doi: https://pubs.acs.org/doi/abs/10.1021/acssuschemeng.8b06374.
[35] Hassanpouryouzband A, Yang J, Tohidi B, Chuvilin E, Istomin V, Bukhanov B, et al. CO2 Capture by Injection of Flue Gas or CO2–N2 Mixtures into Hydrate Reservoirs: Dependence of CO2 Capture Efficiency on Gas Hydrate Reservoir Conditions. Environmental Science & Technology 2018;52(7):4324-30. doi: https://doi.org/10.1021/acs.est.7b05784.
[36] Hassanpouryouzband A, Yang J, Tohidi B, Chuvilin E, Istomin V, Bukhanov B, et al. Insights into CO2 capture by flue gas hydrate formation: gas composition evolution in systems containing gas hydrates and gas mixtures at stable pressures. ACS Sustainable Chemistry and Engineering 2018;6(5):5732-6. doi: https://pubs.acs.org/doi/abs/10.1021/acssuschemeng.8b00409.
[37] Ranaee E, Ghorbani H, Keshavarzian S, Ghazaeipour Abarghoei P, Riva M, Inzoli F, et al. Analysis of the performance of a crude-oil desalting system based on historical data. Fuel 2021;291:120046. doi: https://doi.org/10.1016/j.fuel.2020.120046.
[38] Rashidi S, Mehrad M, Ghorbani H, Wood DA, Mohamadian N, Moghadasi J, et al. Determination of bubble point pressure & oil formation volume factor of crude oils applying multiple hidden layers extreme learning machine algorithms. Journal of Petroleum Science and Engineering 2021:108425. doi: https://doi.org/10.1016/j.petrol.2021.108425.

[39] Seyyedattar M, Ghiasi MM, Zendehboudi S, Butt S. Determination of bubble point pressure and oil formation volume factor: Extra trees compared with LSSVM-CSA hybrid and ANFIS models. Fuel 2020;269:116834. doi: https://doi.org/10.1016/j.fuel.2019.116834.

[40] Daneshfar R, Keivanimehr F, Mohammadi-Khanaposhtani M, Baghban A. A neural computing strategy to estimate dew-point pressure of gas condensate reservoirs. Petroleum Science and Technology 2020;38(10):706-12. doi: https://doi.org/10.1080/10916466.2020.1780257.

[41] Wang K, Luo J, Wei Y, Wu K, Li J, Chen Z. Practical application of machine learning on fast phase equilibrium calculations in compositional reservoir simulations. Journal of Computational Physics 2020;401:109013. doi: https://doi.org/10.1016/j.jcp.2019.109013.

[42] Rashidi S, Khajehesfandeari M. Committee machine-ensemble as a general paradigm for accurate prediction of bubble point pressure of crude oil. Journal of Energy Resources Technology 2021;143(2). doi: https://doi.org/10.1115/1.4047977.

[43] Farsi M, Barjouei HS, Wood DA, Ghorbani H, Mohamadian N, Davoodi S, et al. Prediction of oil flow rate through orifice flow meters: Optimized machine-learning techniques. Measurement 2021;174:108943. doi: https://doi.org/10.1016/j.measurement.2020.108943.

[44] Fadaei M, Ameli F, Hashemabadi SH. Investigation on different scenarios of two-phase flow measurement using Orifice and Coriolis flow meters: Experimental and modeling approaches. Measurement 2021;175:108986. doi: https://doi.org/10.1016/j.measurement.2021.108986.

[45] Dayev ZA. Application of artificial neural networks instead of the orifice plate discharge coefficient. Flow Measurement and Instrumentation 2020;71:101674. doi: https://doi.org/10.1016/j.flowmeasinst.2019.101674.

[46] Jamei M, Ahmadianfar I, Chu X, Yaseen ZM. Estimation of triangular side orifice discharge coefficient under a free flow condition using data-driven models. Flow Measurement and Instrumentation 2021;77:101878. doi: https://doi.org/10.1016/j.flowmeasinst.2020.101878.

[47] Shojaei Barjouei H, Ghorbani H, Mohamadian N, Wood DA, Davoodi S, Moghadasi J, et al. Prediction performance advantages of deep machine learning algorithms for two-phase flow rates through wellhead chokes. Journal of Petroleum Exploration and Production 2021:1-29. doi: https://link.springer.com/article/10.1007/s13202-021-01087-4.

[48] Choubineh A, Ghorbani H, Wood DA, Moosavi SR, Khalafi E, Sadatshojaei E. Improved predictions of wellhead choke liquid critical-flow rates: modelling based on hybrid neural network training learning based optimization. Fuel 2017;207:547-60. doi: https://doi.org/10.1016/j.fuel.2017.06.131.

[49] Ghorbani H, Wood DA, Mohamadian N, Rashidi S, Davoodi S, Soleimanian A, et al. Adaptive neuro-fuzzy algorithm applied to predict and control multi-phase flow rates through wellhead chokes. Flow Measurement and Instrumentation 2020;76:101849. doi: https://doi.org/10.1016/j.flowmeasinst.2020.101849.

[50] Ghorbani H, Wood DA, Moghadasi J, Choubineh A, Abdizadeh P, Mohamadian N. Predicting liquid flow-rate performance through wellhead chokes with genetic and solver optimizers: an oil field case study. Journal of Petroleum Exploration and Production Technology 2019;9(2):1355-73. doi: https://link.springer.com/article/10.1007/s13202-018-0532-6.

[51] Khan MR, Tariq Z, Abdulraheem A. Application of artificial intelligence to estimate oil flow rate in gas-lift wells. Natural Resources Research 2020;29(6):4017-29. doi: https://link.springer.com/article/10.1007/s11053-020-09675-7.

[52] Alakeely AA, Horne RN. Application of deep learning methods to estimate multiphase flow rate in producing wells using surface measurements. Journal of Petroleum Science and Engineering 2021:108936. doi: https://doi.org/10.1016/j.petrol.2021.108936.

[53] Gomaa I, Gowida A, Elkatatny S, Abdulraheem A. The prediction of wellhead pressure for multiphase flow of vertical wells using artificial neural networks. Arabian Journal of Geosciences 2021;14(9):1-10. doi: https://link.springer.com/article/10.1007/s12517-021-07099-y.

[54] Mirzaei-Paiaman A, Salavati S. The application of artificial neural networks for the prediction of oil production flow rate. Energy Sources, Part A: Recovery, utilization, and environmental effects 2012;34(19):1834-43. doi: https://doi.org/10.1080/15567036.2010.492386.

[55] Nejatian I, Kanani M, Arabloo M, Bahadori A, Zendehboudi S. Prediction of natural gas flow through chokes using support vector machine algorithm. Journal of Natural Gas Science and Engineering 2014;18:155-63. doi: https://doi.org/10.1016/j.jngse.2014.02.008.

[56] Gorjaei RG, Songolzadeh R, Torkaman M, Safari M, Zargar G. A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes. Journal of Natural Gas Science and Engineering 2015;24:228-37. doi: https://doi.org/10.1016/j.jngse.2015.03.013.

[57] Rostami A, Ebadi H. Toward gene expression programming for accurate prognostication of the critical oil flow rate through the choke: Correlation development. Asia-Pacific Journal of Chemical Engineering 2017;12(6):884-93. doi: https://doi.org/10.1002/apj.2126.

[58] ZareNezhad B, Aminian A. Accurate prediction of the dew points of acidic combustion gases by using an artificial neural network model. Energy conversion management 2011;52(2):911-6. doi: https://doi.org/10.1016/j.enconman.2010.08.018.

[59] Elhaj MA, Anifowose F, Abdulraheem A. Single gas flow prediction through chokes using artificial intelligence techniques. *SPE Saudi Arabia Section Annual Technical Symposium and Exhibition.* Society of Petroleum Engineers; 2015. doi: https://doi.org/10.2118/177991-MS.

[60] Yuan X, Chen C, Yuan Y, Huang Y, Tan Q. Short-term wind power prediction based on LSSVM–GSA model. Energy Conversion and Management 2015;101:393-401. doi: https://doi.org/10.1016/j.enconman.2015.05.065.

[61] Ahmadi MA, Ebadi M, Hosseini SM. Prediction breakthrough time of water coning in the fractured reservoirs by implementing low parameter support vector machine approach. Fuel 2014;117:579-89. doi: https://doi.org/10.1016/j.fuel.2013.09.071.

[62] Ahmadi M-A, Bahadori A. A LSSVM approach for determining well placement and conning phenomena in horizontal wells. Fuel 2015;153:276-83. doi: https://doi.org/10.1016/j.fuel.2015.02.094.

[63] Farsi M, Mohamadian N, Ghorbani H, Wood DA, Davoodi S, Moghadasi J, et al. Predicting Formation Pore-Pressure from Well-Log Data with Hybrid Machine-Learning Optimization Algorithms. Natural Resources Research 2021:1-27. doi: https://link.springer.com/article/10.1007/s11053-021-09852-2.

[64] Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural processing letters 1999;9(3):293-300. doi: https://link.springer.com/article/10.1023/A:1018628609742.
[65] Suykens JA, Van Gestel T, De Brabanter J. Least squares support vector machines. World scientific; 2002.
[66] Ahmadi MA, Ebadi M. Evolving smart approach for determination dew point pressure through condensate gas reservoirs. Fuel 2014;117:1074-84. doi: https://doi.org/10.1016/j.fuel.2013.10.010.
[67] Eslamimanesh A, Gharagheizi F, Mohammadi AH, Richon D. Phase equilibrium modeling of structure H clathrate hydrates of methane+ water “insoluble” hydrocarbon promoter using QSPR molecular approach. Journal of Chemical & Engineering Data 2011;56(10):3775-93. doi: https://doi.org/10.1021/je200444f.
[68] Ahmadi MA, Zahedzadeh M, Shadizadeh SR, Abbassi R. Connectionist model for predicting minimum gas miscibility pressure: Application to gas injection process. Fuel 2015;148:202-11. doi: https://doi.org/10.1016/j.fuel.2015.01.044.
[69] Ahmadi MA. Connectionist approach estimates gas–oil relative permeability in petroleum reservoirs: application to reservoir simulation. Fuel 2015;140:429-39. doi: https://doi.org/10.1016/j.fuel.2014.09.058.
[70] Ahmadi MA, Mahmoudi B. Development of robust model to estimate gas–oil interfacial tension using least square support vector machine: experimental and modeling study. The Journal of Supercritical Fluids 2016;107:122-8. doi: https://doi.org/10.1016/j.supflu.2015.08.012.
[71] Ahmadi MA. Toward reliable model for prediction drilling fluid density at wellbore conditions: A LSSVM model. Neurocomputing 2016;211:143-9. doi: https://doi.org/10.1016/j.neucom.2016.01.106.
[72] Ahmadi MA, Pournik M. A predictive model of chemical flooding for enhanced oil recovery purposes: Application of least square support vector machine. Petroleum 2016;2(2):177-82. doi: https://doi.org/10.1016/j.petlm.2015.10.002.
[73] Ahmadi MA, Rozyn J, Lee M, Bahadori A. Estimation of the silica solubility in the superheated steam using LSSVM modeling approach. Environmental Progress & Sustainable Energy 2016;35(2):596-602. doi: https://doi.org/10.1002/ep.12251.
[74] Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: theory and applications. Neurocomputing 2006;70(1-3):489-501. doi: https://doi.org/10.1016/j.neucom.2005.12.126.
[75] Huang G-B, Siew C-K. Extreme learning machine with randomly assigned RBF kernels. International Journal of Information Technology 2005;11(1):16-24. doi: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.1555&rep=rep1&type=pdf.
[76] Jiang X, Yan T, Zhu J, He B, Li W, Du H, et al. Densely Connected Deep Extreme Learning Machine Algorithm. Cognitive Computation 2020;12(5):979-90. doi: https://link.springer.com/article/10.1007/s12559-020-09752-2.
[77] Ahmadi MA, Ebadi M, Shokrollahi A, Majidi SMJ. Evolving artificial neural network and imperialist competitive algorithm for prediction oil flow rate of the reservoir. Applied Soft Computing 2013;13(2):1085-98. doi: https://doi.org/10.1016/j.asoc.2012.10.009.

[78] Ahmadi MA, Chen Z. Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. Petroleum 2019;5(3):271-84. doi: https://doi.org/10.1016/j.petlm.2018.06.002.
[79] Sajjadi S, Shamshirband S, Alizamir M, Yee L, Mansor Z, Manaf AA, et al. Extreme learning machine for prediction of heat load in district heating systems. Energy and Buildings 2016;122:222-7. doi: https://doi.org/10.1016/j.enbuild.2016.04.021.
[80] Tang J, Deng C, Huang G-B, Hou J. A fast learning algorithm for multi-layer extreme learning machine. *2014 IEEE International Conference on Image Processing (ICIP).* IEEE; 2014:175-8. doi: https://doi.org/10.1109/ICIP.2014.7025034.
[81] Mokarizadeh H, Atashrouz S, Mirshekar H, Hemmati-Sarapardeh A, Pour AM. Comparison of LSSVM model results with artificial neural network model for determination of the solubility of SO2 in ionic liquids. Journal of Molecular Liquids 2020;304:112771. doi: https://doi.org/10.1016/j.molliq.2020.112771.
[82] Mo L, Chen H, Chen W, Feng Q, Xu L. Study on evolution methods for the optimization of machine learning models based on FT-NIR spectroscopy. Infrared Physics & Technology 2020;108:103366. doi: https://doi.org/10.1016/j.infrared.2020.103366.
[83] Shavinina LV. The international handbook on innovation. Elsevier; 2003.
[84] Goldberg DE, Holland JH. Genetic algorithms and machine learning. 1988. doi: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/46947/10994_2005_Article_422926.pdf.
[85] Sivanandam S, Deepa S. Genetic algorithms. Introduction to genetic algorithms. Springer; 2008, p. 15-37.
[86] Ahmadi M, Chen Z. Machine learning-based models for predicting permeability impairment due to scale deposition. Journal of Petroleum Exploration and Production Technology 2020;10(7):2873-84. doi: https://link.springer.com/article/10.1007/s13202-020-00941-1.
[87] Eberhart R, Kennedy J. A new optimizer using particle swarm theory. *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science.* Ieee; 1995:39-43. doi: https://doi.org/10.1109/MHS.1995.494215.
[88] Kuo R, Hong S, Huang Y. Integration of particle swarm optimization-based fuzzy neural network and artificial neural network for supplier selection. Applied Mathematical Modelling 2010;34(12):3976-90. doi: https://doi.org/10.1016/j.apm.2010.03.033.
[89] Kıran MS, Özceylan E, Gündüz M, Paksoy T. A novel hybrid approach based on particle swarm optimization and ant colony algorithm to forecast energy demand of Turkey. Energy conversion and management 2012;53(1):75-83. doi: https://doi.org/10.1016/j.enconman.2011.08.004.
[90] Katiyar S. A comparative study of genetic algorithm and the particle swarm optimization. International Journal of Technology 2010;2(2):21-4. doi:
[91] Hassan R, Cohanim B, De Weck O, Venter G. A comparison of particle swarm optimization and the genetic algorithm.1897. doi: https://doi.org/10.2514/6.2005-1897.
[92] Duan Y, Harley RG, Habetler TG. Comparison of particle swarm optimization and genetic algorithm in the design of permanent magnet motors. IEEE:822-5. doi: 10.1109/IPEMC.2009.5157497.

[93] Panda S, Padhy NP. Comparison of particle swarm optimization and genetic algorithm for FACTS-based controller design. Applied soft computing 2008;8(4):1418-27. doi: https://doi.org/10.1016/j.asoc.2007.10.009.

[94] Hodge V, Austin J. A survey of outlier detection methodologies. Artificial intelligence review 2004;22(2):85-126. doi: https://link.springer.com/article/10.1023/B:AIRE.0000045502.10941.a9.

[95] Kovács L, Vass D, Vidács A. Improving quality of service parameter prediction with preliminary outlier detection and elimination. *Proceedings of the second international workshop on inter-domain performance and simulation (IPS 2004), Budapest.* 2004. 2004:194-9. doi:

[96] Mehrad M, Bajolvand M, Ramezanzadeh A, Neycharan JG. Developing a new rigorous drilling rate prediction model using a machine learning technique. Journal of Petroleum Science and Engineering 2020;192:107338. doi: https://doi.org/10.1016/j.petrol.2020.107338.

[97] Myers L, Sirois MJ. S pearman correlation coefficients, differences between. Encyclopedia of statistical sciences 2004. doi: https://doi.org/10.1002/0471667196.ess5050.

[98] Artusi R, Verderio P, Marubini E. Bravais-Pearson and Spearman correlation coefficients: meaning, test of hypothesis and confidence interval. The International journal of biological markers 2002;17(2):148-51. doi: https://doi.org/10.1177/172460080201700213.