


Original article

Application of mathematical and machine learning models to predict differential pressure of autonomous downhole inflow control devices

Hossein Yavari¹, Rasool Khosravanian², David A. Wood³^{*}, Bernt Sigve Aadnoy²

¹Department of Petroleum Engineering, Amirkabir University of Technology, Tehran, Iran

²Department of Energy and Petroleum Engineering, University of Stavanger, Stavanger, Norway

³DWA Energy Limited, Lincoln, United Kingdom

Keywords:

Autonomous inflow control
downhole differential pressure
intelligent non-linear models
multiple regression
machine learning
optimum viscosity

Cited as:

Yavari, H., Khosravanian, R., Wood, D. A., Aadnoy, B. S. Application of mathematical and machine learning models to predict differential pressure of autonomous downhole inflow control devices. *Advances in Geo-Energy Research*, 2021, 5(4): 386-406, doi: 10.46690/ager.2021.04.05

Abstract:

Controlling reservoir fluid flow is important for maximizing petroleum production through wellbores. A major challenge that reduces the production of oil is early breakthrough of secondary fluids to the wellbore perforations. This occurs due to the low viscosity of gas and water relative to oil, and the heterogeneity of reservoir permeability. Autonomous inflow control devices represent a new self-regulating technology that helps to increase petroleum production, particularly oil, by restricting the production of unwanted fluids like gas and water into the wellbores. This study develops smart systems based on machine learning models to predict the performance of autonomous inflow control devices. Several machine learning models are evaluated including adaptive neuro fuzzy inference system, hybrid adaptive neuro-fuzzy inference system genetic algorithm, artificial neural network and support vector machine and their prediction performance is compared to that of linear regression, full quadratic regression model and the mathematical autonomous inflow control device performance model. Each model is developed to estimate the differential pressure of Equiflow autonomous inflow control devices based on ninety experimentally recorded data records. The range of equiflow autonomous inflow control device, viscosity, density and flow rate are the input variables and differential pressure is the output dependent variable of each model. The prediction accuracy of the models is assessed in terms of several standard statistical accuracy performance measures. These performance indicators confirm that the machine-learning models provide superior prediction accuracy for autonomous inflow control device differential pressure. Overall, the support vector machine achieves the most accurate predictions of all the models evaluated recording root mean square error of 0.14 Mpa and coefficient of determination of 0.98. On the other hand, the linear regression model records the lowest prediction performance, highlighting the non-linearity of the autonomous inflow control device processes.

1. Introduction

Well completion technology has improved remarkably during the past three decades (Eltaher, 2017). “Smart” well completion technologies, many providing real time data recording, are being developed for oil and gas field deployments around the world (Glandt, 2005). Shutting-in zones providing poor production performance, remotely back and monitoring the production of wells are some of the capabilities of these smart technologies (Jovanov, 2016). The heterogeneous na-

ture of reservoirs causes several production-related problems, including viscous fingering of reservoir fluid movements and production of unwanted fluids, such as water and gas, especially in long horizontal wells (Chengzao, 2017; Zhao et al., 2019; Cheng et al., 2020). Advanced well completions utilize downhole flow control technology like interval control valves, inflow control devices (ICD) and autonomous inflow control devices (AICD) to provide practical solutions to the production problems and constraints (Gimre, 2012; Eltaher, 2017).

Maximizing oil recovery from subsurface reservoirs is a

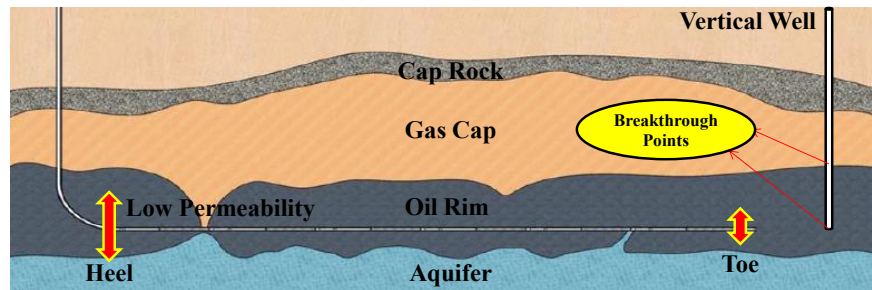


Fig. 1. Schematic of fluid breakthrough in horizontal and vertical wells (modified after Aakre et al., 2013).

massive challenge for the oil production sector. Increasing the reservoir contact per well and decreasing the negative effects of water and gas breakthrough are the most important factors to consider when striving to increase oil recovery from a reservoir and enhance oil production rate. Traditionally, vertical wells were used to access oil reservoirs. Although the drilling of vertical wells is a simpler technique, such wells have limited reservoir contact (Mikkelsen et al., 2005; Aakre et al., 2013).

Fig. 1 reveals why better reservoir contact is achieved by drilling deviated wells, especially when they include long horizontal or low-angle sections. In horizontal wells early breakthrough of water and/or gas is prone to occur in both homogeneous and heterogeneous reservoirs. In homogeneous reservoirs the fluid friction is the only reason causing pressure drop within the wellbore. That pressure drop is proportional to the density, viscosity, flow rate, length and diameter of the well. Hence, pressure in the toe of the well is greater than pressure in the heel of the well. As pressure drop in the heel of the well increases, it leads to more oil production in the heel of the well, causing early breakthrough of gas and/or water to occur in the heels of horizontal wells (Al-Khelaiwi and Davies, 2007; Aakre et al., 2013).

Heterogeneities, for example permeability differences along the wellbore, may also result in early gas/oil breakthrough. In these situations well logs are used to determine the most susceptible zones with high permeability. Such zones need to be avoided to delay early gas/oil breakthrough or decrease the negative effects of such breakthrough by back unwanted fluid production. Therefore utilizing ICDs and AICDs in horizontal wells is necessary to diminish the negative effects of early gas/oil breakthrough. In vertical wells, fluid breakthrough tends to occur at the highest and lowest perforations, because they are closer to gas cap (if present) and aquifer, respectively (Aakre et al., 2013).

Traditional ICD, controlled only from the surface, have been used as production aids since the 1990's (Eltaher, 2017). However, this technology generally proved to be too expensive due to it requiring an energy source. These constraints meant that only a limited number of them (typically four to five ICDs) could be installed in single well. Consequently, most operating companies abandoned this early device due to its high cost and low reliability (Gao et al., 2007; Halvorsen et al., 2012). Recently, hydraulic control systems have substantially improved, leading to more extensive uptake and reliability of

down-hole control devices.

AICD are a new technology that has stimulated industry interest due to their low costs, simplicity, precision, autonomous operation, and, because they do not require an energy source (Aakre et al., 2013). The performance of these devices is controlled by viscosity differences of fluids and fluid mechanics principles, including Bernoulli's principle, and static and dynamic fluid pressure variations through the valve, making them powerful and efficient operating devices (Aakre et al., 2014; Zhao et al., 2014; Jovanov, 2016; Eltaher, 2017). Also, unlike ICDs, a large number of them can be used in one well (Halvorsen et al., 2012).

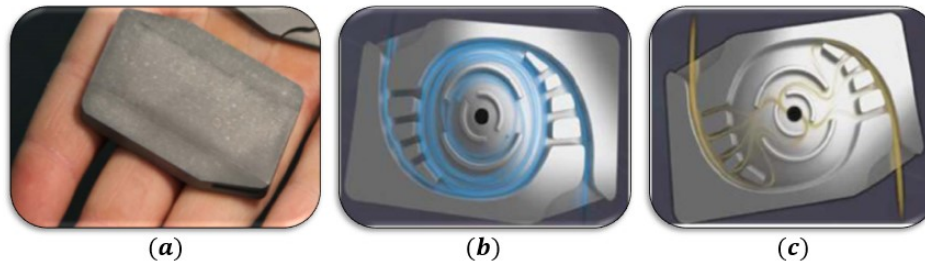
After breakthrough, AICDs significantly reduce unwanted fluid production (e. g., water and gas), which increases oil production and also reduces costs of separation and re-injection of unwanted fluids produced along with oil to the surface (Aakre et al., 2013; Andreas et al., 2019; Lei et al., 2020). Cochua et al. (2018) and Gurses et al. (2019) introduced a cyclone type AICD which is similar to Equiflow AICD and performs based on the viscosity and density of flowing fluids. Zhang et al. (2019) designed a new Cyclone Inflow Control Device using the principle of the cyclone to deal with fine particles like sand. Equiflow is a self-regulating inflow control device with no moving parts that is based on the fluidic diode principle (Fripp et al., 2013), i.e., it is a passive device that displays reduced flow resistance in its preferred direction of flow compared to counter directions. The performance of the valve is connected with the properties of the producing fluid including viscosity, density and flow rate (Jovanov, 2016). An Equiflow AICD comes in four different ranges which are tailored to a range of oil viscosities from very light oil to very heavy oil (Eltaher, 2017). Table 1 displays details of the four versions and their properties.

Fig. 2 illustrates the general structure and size of Equiflow AICDs and the fluid flow-pathways for high (such as heavy oil) and low (like water) viscous fluids (Pedroso et al., 2020).

Fluids with low viscosity, including gas and water, take the longer path and begin to spin in the vortex of the device (Fig. 2). As a result, pressure drop increases and this, in turn, further acts to choke the gas and water production. Due to the effect of viscous forces, viscous fluids take a more direct pathway with lower pressure drop (Jovanov, 2016). Therefore, when gas or water breakthrough occurs, the Equiflow AICD significantly restricts unwanted fluid production from that specific section while promoting increased oil production from

Table 1. Equiflow AICD specialized designs (Eltaher, 2017).

No	Design	Oil viscosity range (cp)	Oil type	Fluid restrictions
1	Range 1	0.3 - 1.5	Very light	Water and Gas
2	Range 2	1.5 - 10	Light, medium	Water and Gas
3	Range 3	3 - 200	Light, medium heavy	Water and Gas
4	Range 4	150 +	Heavy, very heavy	Water and Gas

**Fig. 2.** Equiflow AICD (a) size, (b) water-flow pathway (right) and (c) oil-flow pathway (left) (Fripp et al., 2013; Iqbal et al., 2015).

other compartments in the well completion (Least et al., 2012; Zhao et al., 2014; Porturas, 2016).

There is an optimum fluid viscosity for which production is favored, for each Equiflow AICD. The minimum differential pressure occurs at the optimum viscosity and beyond the optimum viscosity the differential pressure increases. As a result, the performance of the Equiflow AICD changes with changing fluid viscosity and, in certain conditions and fluid compositions this can act to restrict the production of oil (Least et al., 2012).

To meaningfully predict an AICD's flow performance it is necessary to develop a model that is able to estimate the differential pressure across the AICD accurately using a small data set. Unfortunately, in field conditions it is not possible to record numerous data measurements of viscosity and density of the reservoir fluid. To do so is both technically difficult time consuming and costly. Moreover, reservoir density and viscosity vary across limited ranges. Therefore, a practical method must be able to model the performance of the Equiflow AICDs accurately in field conditions based upon a limited set of data. Accurate AICD prediction models are beneficial for reservoir simulation, assisting them to accurately predict the production rate and efficiency of reservoirs on a field-wide basis (Eltaher, 2017; Pedroso et al., 2020). Conventional autonomous flow control device (AFCD) formula, Halliburton model, Aadnoy model and the mathematical AICD performance model are the most popular mathematical models used to estimate the differential pressure of AICDs (Jovanov, 2016; Eltaher, 2017; Eltaher et al., 2019). However, results show that these mathematical models are not able to model the optimum performance of AICDs accurately. In these models, for a constant differential pressure, as viscosity increases the flow rate also increases. However, this is incorrect, as outside the optimum viscosity range the flow rate actually decreases in these devices. On the other hand, mathematical models can only be developed for one valve, and when

there are several valves in a well, a separate model must be developed for each valve. In recent years, many soft computing techniques and machine-learning methods, some hybridized with efficient optimization algorithms, have been adopted as powerful approaches to predict various parameters associated with complex systems in the oil and gas industry (Rabiei et al., 2015; Jovic et al., 2016; Wood, 2018; Yavari et al., 2018; Ashfari et al., 2019; Barbosa et al., 2019; Rashid et al., 2019; Sabah et al., 2019; Yilmaz et al., 2019; Elkatatny, 2020; Gamal et al., 2020; Ghorbani et al., 2020; Mehrad et al., 2020; Moazzeni and Khamehchi, 2020; Ossai and Duru, 2020; Somehsaraei et al., 2020; Abad et al., 2021; Hazbeh et al., 2021; Mardanirad et al., 2021; Mohamadian et al., 2021).

The aim of this study is to develop an accurate and reliable model to estimate the performance of Equiflow AICDs and to evaluate various machine learning and multiple regression models that could potentially achieve this. As there is an optimum viscosity for each specified Equiflow AICD, successful models must be able to identify the optimum performance of Equiflow AICDs accurately under various downhole conditions and fluid combinations, i.e., oil, gas and water cuts. Several studies have been performed that simulate the effects of both ICDs and AICDs on oil recovery, using commercial software including OLGA, ROCX, Petrel and Eclipse. Mathematical models are combined with such software to calculate the pressure drop associated with smart valves. However, the mathematical models used by the software typically do not consider the optimum viscosity of these valves. Hence, the results of this study could be used to increase the accuracy of such simulations (Mikkelsen et al., 2005, Aakre et al., 2013; Eltaher, 2017; Eltaher et al., 2019).

2. Data collection

Least et al. (2012) performed laboratory tests to assess the performance of range 3 and range 4 Equiflow AICDs in field-like conditions. The designed test set up includes a

triplex pump, tank, pipes, and a circulating loop. A shell and tube heat exchanger which is supplied with chilled and heated water is utilized to control the temperature of the fluid in the system. Most of the flow lines are also insulated to assist with temperature control. The smart valve is placed on the horizontally-oriented pipe and the system is configured to measure the static pressure, flow rate and temperature at the inlet and outlet of the AICD. The viscosity of the flowing fluid is measured using a laboratory viscometer, set up in accordance with the ASTM D-445 standard, and the density of the fluid is determined based on the ASTM D-4052 standard, in field-like conditions (Least et al., 2012). Zhao et al. (2014) also performed laboratory tests to measure the performance of the range 2 Equiflow AICD simulating field like conditions. Liquid testing required a two-phase-flow facility, specifically designed to assess oil and water mixtures. Gas testing required a nitrogen-gas-blowdown facility. For testing liquids, an oil-water separator acted as a two-phase tank with a single triplex pump circulating the test fluids (oil, water) in the flow loop. During testing, fluid was drawn through either the oil suction or water suction exits from the separator. A Coriolis flow meter was used to provide the reference liquid flow rate and density prior to entry into the AICD. A flowthrough circulation heater provided test-fluid-temperature control. The AICD Range 2 utilizes similar fluid vectoring to the AICD Range 3 device. However, the range 2 device includes a more autonomous on/off type switching function whereas the Range 3 operates

in response to gradual changes in performance. Tests show that the AICD range 2 is most effective with oil viscosities in the 1.5-10 cP range (Zhao et al., 2014).

The viscosity, density and flow rate are the most important properties of flowing fluids on which the design and performance of the Equiflow AICD are based. Therefore, in all models the viscosity, density, flow rate and range of the valves are the input variables considered and the differential pressure across the device is the output / dependent variable. 90 data points are compiled from published evaluations of ranges 2, 3 and 4 Equiflow AICD (Least et al., 2012; Zhao et al., 2014). The data set is divided into two parts, in which 80 % of overall data are used to develop the model (training subset), and the remaining 20 % are used to test the developed model (testing subset). Table 2 shows detailed measured variable values for each of 90 individual data records divided into training (72 records, 80 % of full dataset) and testing (18 records, 20 % of full dataset) subsets.

Table 3 displays the statistical details of the data set.

Fig. 3 graphically displays the differential pressure versus flow rate of flowing fluids with various viscosities for range 3 Equiflow AICDs. It illustrates that by increasing the flow rates of the low-viscosity fluids, such as water, significantly increases the differential pressure and reduces their production. These relationships are clearly non-linear. Increasing fluid viscosity reduces differential pressure, which continues until a viscosity of 0.099. Further increasing the viscosity to 0.229,

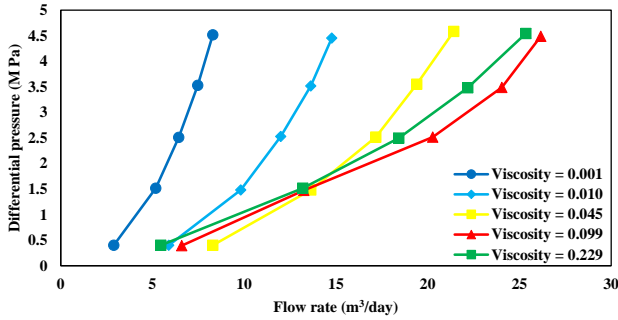
Table 2. Equiflow AICDs performance data (Least et al., 2012; Zhao et al., 2014).

No.	Independent variables				Dependent variable
	Range	Viscosity (Kg/m-s)	Density (S.G.)	Flow rate (m ³ /day)	Differential pressure (MPa)
33	3	0.001	0.994	2.889	0.400
34	3	0.001	0.994	6.432	2.510
35	3	0.001	0.994	7.468	3.530
36	3	0.001	0.994	8.286	4.516
37	3	0.010	0.849	5.887	0.400
38	3	0.010	0.849	9.812	1.482
39	3	0.010	0.849	11.992	2.530
40	3	0.010	0.849	14.772	4.454
41	3	0.045	0.849	8.286	0.400
42	3	0.045	0.849	17.171	2.517
43	3	0.045	0.849	19.406	3.551
44	3	0.045	0.849	21.422	4.585
45	3	0.099	0.865	6.596	0.393
46	3	0.099	0.865	13.246	1.475
47	3	0.099	0.865	20.278	2.517
48	3	0.099	0.865	26.165	4.488
49	3	0.229	0.881	5.451	0.400
50	3	0.229	0.881	18.424	2.496
51	3	0.229	0.881	22.186	3.482
52	3	0.229	0.881	25.347	4.544

Only the AICD Range 3 training records are shown. For all 90 data records see the Supplementary file.

Table 3. Statistical parameters of input and output data.

No.	Parameter	Unit	Minimum	Maximum	Average
1	Viscosity	Kg/m-s	0.0002	1.002	0.5011
2	Density	S.G.	0.79	0.995	0.8925
3	Flow rate	m ³ /day	2.235	31.18	16.7075
4	Range	-	2	4	3
5	Differential pressure	MPa	0.109	4.585	2.347

**Fig. 3.** Range 3 Equiflow AICDs differential pressure versus flow rate performance curves (Least et al., 2012; Zhao et al., 2014).

however, increases the differential pressure again. It reveals that the optimum viscosity of the range 3 Equiflow AICD for any specific flow rate is between 0.099 to 0.229 Kg/m-s.

3. Mathematical model

Mathiesen et al. (2011) developed a function for Rate Controlled Production valves to estimate the differential pressure according to several experimental data. Eq. (2), which is known as the conventional AFCD formula, is derived based on a classical dimensional analysis for modeling the performance of AICDs. This new formula has better prediction performance than the other formulas applied, such as Halliburton's method. In particular, it provides better accuracy and physical representation of multiphase-flow performance (Eltaher, 2017). This formula is expressed as Eqs. (1) to (3) (Lauritzen and Martiniussen, 2011; Mathiesen et al., 2011; Eltaher, 2017):

$$\delta P = \left(\frac{\rho_{mix}^{x-1}}{\mu_{mix}^{x-2}} \right) b_{AICD} q^x \quad (1)$$

$$\rho_{mix} = (\alpha_o)^a \rho_o + (\alpha_w)^b \rho_w + (\alpha_g)^c \rho_g \quad (2)$$

$$\mu_{mix} = (\alpha_o)^d \mu_o + (\alpha_w)^e \mu_w + (\alpha_g)^f \mu_g \quad (3)$$

where x is volume flow-rate exponent, b_{AICD} is a constant referred to as "strength of AICD" that has a dimension of $(length)^{-(x+2)}$, $x-2$ is the viscosity function's exponent, $x-1$ is the density function's exponent, μ_{mix} and ρ_{mix} are volumetric averages of fluid viscosity and fluid density, respectively, that are defined by Eqs. (2) and (3). The term α is the volume fraction of each phase and the parameters a, b, c, d, e and f are mixture exponents each of which are usually assumed to

equal one (Eltaher, 2017). The subscripts o , w and g represent oil, water and gas, respectively.

3.1 Mathematical model development

A genetic algorithm (GA) assists in the determination of constant coefficients for the mathematical models based on the available experimental data. GA was introduced by John Holland in the 1960s using analogies to biological evolution concepts (Goldberg and Holland, 1988; Holland, 1992a; Mitchell, 1996). GA conducts its searches of the feasible solution spaces using metaheuristics that emulate natural selection and other genetic processes. GA commences with a randomly generated population of solutions subject to defined limitations and constraints. At the end of each iteration the set of solutions is ranked in terms of its prediction performance of the dependent variable's values for the population as a whole and the attributes of the population amended accordingly. The set of solutions are called population and each solution is called a chromosome. Individual solutions (analogous to chromosomes) evolves through a series of iterations (analogous to generations). The next generation of individual solutions (analogous to offspring) are created in several ways: (1) by combining the attributes of two high-performing individual solutions from the current iteration applying a crossover operator, several of which are widely used; 2) applying a mutation operator to certain attributes of high performing solutions from the current iteration; and, 3) generating a small, specified number of new individual solutions randomly. Each individual solution is evaluated by a defined measure of fitness (minimizing an objective function) during each iteration to ensure that only the "fitter" high performing individual solutions are either selected to progress to the next iteration, or are involved in the crossover and mutation modifications to create new solutions for the next iteration. This elitist strategy helps the GA algorithm to progress steadily to optimum solutions through a series of iterations. After several iterations the GA ultimately converges to an optimum solution, although there is a risk that it may become trapped at local optima rather than reach the global optimum depending on the nature of the solution space (Holland, 1992b; Mitchell, 1996; Velez-Langs, 2005; Rahul et al., 2011).

In this study, the differential pressure is the output of all models evaluated. However, in order to develop models and find the optimum constant coefficients of the models using experimental data, the root mean square error (RMSE) is employed as the objective function. The goal is to find

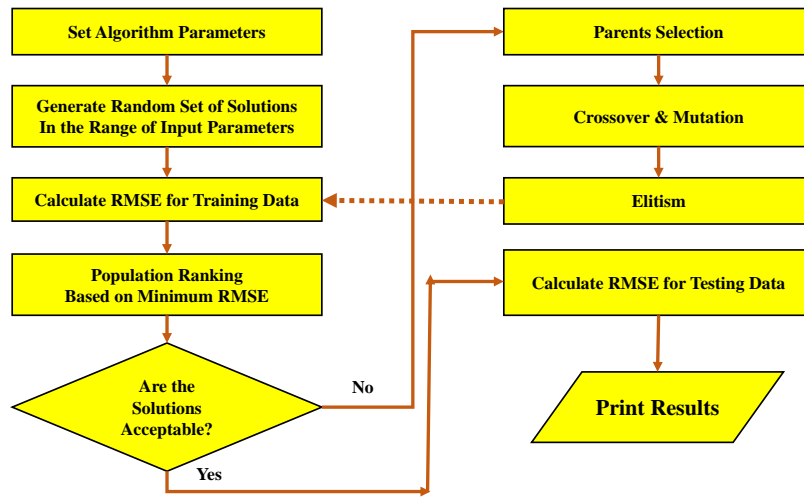


Fig. 4. Flowchart for determining mathematical model coefficients applying a GA.

the constant coefficients of the model that have the least error and provide the closest match with the collection of experimental data records taken as a whole. Therefore, the RMSE is typically considered as the objective function for GA algorithms. The RMSE is calculated using the experimental values of differential pressure (δP_e) and predicted values of differential pressure (δP_p) using the mathematical model under evaluation. RMSE is computed using Eq. (4):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\delta P_e - \delta P_p)^2} \quad (4)$$

where n is the number of data records.

This formula is used in all models to determine the optimum coefficients that have the least error and closest match with the collection of experimental data records taken as a whole.

The flowchart for implementing GA optimization is as shown in Fig. 4.

Table 4 lists the control parameter values applied in this study when implementing the GA.

Table 5 lists the range (lower limit and upper limit) of mathematical model constants coefficients that are applied as constraints for possible GA solutions.

Fig. 5 illustrates GA performance in determining the constant coefficients of the mathematical model for the range 3 Equiflow AICD. It establishes that the determined constant for the AICD performance model (Eq. (1)) has an RMSE value equal to 0.533672 MPa. The lower diagram (Fig. 5) illustrates the best determined values of the constants.

The lowest RMSE values determined for the ranges 2, 3 and range 4 are 0.283627, 0.533672 and 0.33228 MPa, respectively. The tuned (optimum) coefficients for AICD performance model for the ranges 2, 3 and 4 Equiflow AICDs are displayed in Eqs. (5), (6) and (7) respectively, in which b_{AICD} is in $\left[10^9 * 8.64^x (m)^{-(x+2)}\right]$ that is produced for the following units – viscosity (Kg/(m·s)), density (S.G.), flow rate (m³/day) and differential pressure (MPa):

Table 4. Genetic algorithm control parameters.

No	GA parameter	Value
1	Population size	1000
2	Generation	2000
3	Selection method	Roulette wheel
4	Mutation rate	0.8
5	Crossover rate	0.2
6	Elitism rate	0.05 * (Population size)

Table 5. Recommended constraint boundaries for constant coefficients involved in a range 3 AICD mathematical models.

No	Constant coefficient	Lower limit	Upper limit
1	x	0	3
2	b_{AICD}	0	1

$$\delta P = 0.0000340491 \left(\frac{\rho_{mix}^{1.531432}}{\mu_{mix}^{0.531432}} \right) q^{2.531432} \quad \text{Range(2)} \quad (5)$$

$$\delta P = 0.0005393907 \left(\frac{\rho_{mix}^{1.527103}}{\mu_{mix}^{0.527103}} \right) q^{2.527103} \quad \text{Range(3)} \quad (6)$$

$$\delta P = 0.0036429406 \left(\frac{\rho_{mix}^{1.357872}}{\mu_{mix}^{0.357872}} \right) q^{2.357872} \quad \text{Range(4)} \quad (7)$$

4. Multiple regression models

Many engineering problems require the determining of complex relationships between several variables (Khademi et al., 2017). Regression analysis is widely used traditional for such purposes. Linear regression is a special form of regression models that utilizes linear functions to model the

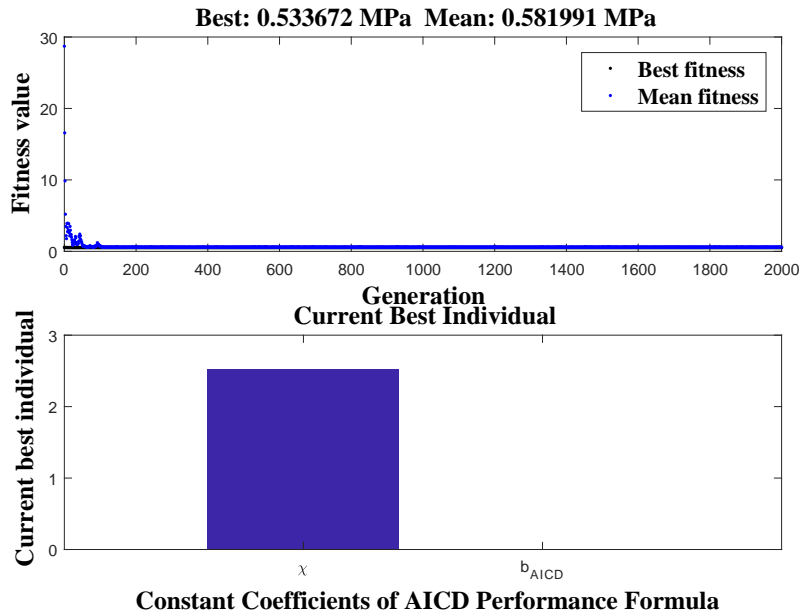


Fig. 5. Performance of genetic algorithm in determining all constant coefficients of the AICD performance model. The upper diagram shows the fitness values (MPa) over multiple GA iterations (RMSE).

relationships among the influencing variables (Nathans et al., 2012; Sadrmomtazi et al., 2013).

4.1 Multiple linear regression

In most systems, there is more than one influencing variable, therefore requiring multiple linear regression (MLR) analysis to predict the dependent variable. MLR attempt to find the relationship between the multiple influencing or independent variables and, usually, a single response or dependent variable by minimizing errors. MLR can be expressed in general terms by Eq. (8) (Sekki et al., 2015; Vickers, 2017; Ciulla and D'Amico, 2019):

$$Y = a_0 + \sum_{j=1}^n a_j X_j \quad (8)$$

where Y is the dependent variable, X_j is the j^{th} independent (input) variable, n is the number of data records, and a represents the partial regression coefficient (Khademi et al., 2017). In this study, GA is used to determine the MLR coefficients. The procedure, GA control-parameter values and training subset sizes are the same as those applied to establish the coefficients of the AICD performance model in the previous section. Viscosity, density and flow rate are the input parameters and differential pressure is the dependent variable.

Fig. 6 displays the performance of GA for determining the constant coefficients of MLR model for range 3 Equiflow AICD. The determined constants have an RMSE of 0.75199 MPa. The lower diagram (Fig. 6) illustrates the determined constants.

The lowest RMSE values achieved for the ranges 2, 3 and range 4 devices are 0.33943, 0.75199 and 0.60063 MPa, respectively. The tuned (optimum) coefficients for the MLR model for ranges 2, 3 and 4 of Equiflow AICDs are provided

in Eqs. (9), (10) and (11), respectively:

$$\delta P = -4.003913 + 14.2574148\mu + 3.94667315\rho + 0.09006775q \quad \text{Range(2)} \quad (9)$$

$$\delta P = -12.02370690 - 6.40040759\mu + 13.37288776\rho + 0.23413315q \quad \text{Range(3)} \quad (10)$$

$$\delta P = -17.43046162 - 0.75110657\mu + 18.64856115\rho + 0.27012850q \quad \text{Range(4)} \quad (11)$$

4.2 Full quadratic multiple regression

A more comprehensive form of multiple regression, more suited to curved relationships between variables, is the full quadratic multiple regression method. The general form of full quadratic multiple regression for dependent variable Y and k independent variables X is expressed by Eq. (12) (Aiken et al., 2012; Moraveji and Naderi, 2016):

$$Y = \beta_0 + \sum_{j=1}^k \beta_j X_j + \sum_{i<j}^{k-1} \sum_{j=2}^k \beta_{ij} X_i X_j + \sum_{j=1}^k \beta_{jj} X_j^2 \quad (12)$$

where β_0 is the model intercept, β_j is linear coefficient, β_{jj} is quadratic coefficient, β_{ij} is interaction coefficient, X_i and X_j are independent variables, Y is the dependent variable of the model and k is the number of independent variables (Moraveji and Naderi, 2016). Formulating this for the AICD model, viscosity, density and flow rate are the input variables, differential pressure is the dependent variable and k is equal to 3. GA is used to determine the constant coefficients of the model. There is no lower and upper bound for constant coefficients.

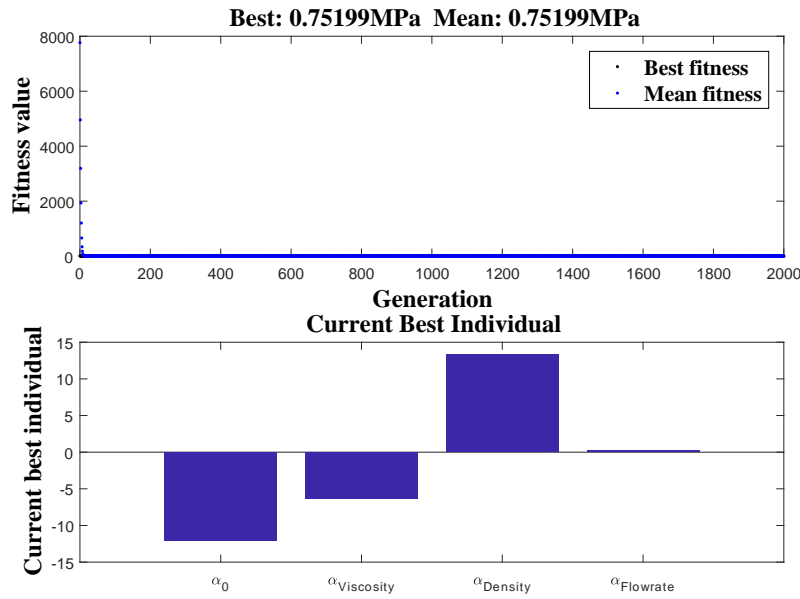


Fig. 6. Performance of the genetic algorithm in determining the four constant coefficients (α_0 , α (viscosity), α (density); α (flow rate)) of MLR model. The upper diagram shows the fitness values over multiple GA iterations (RMSE).

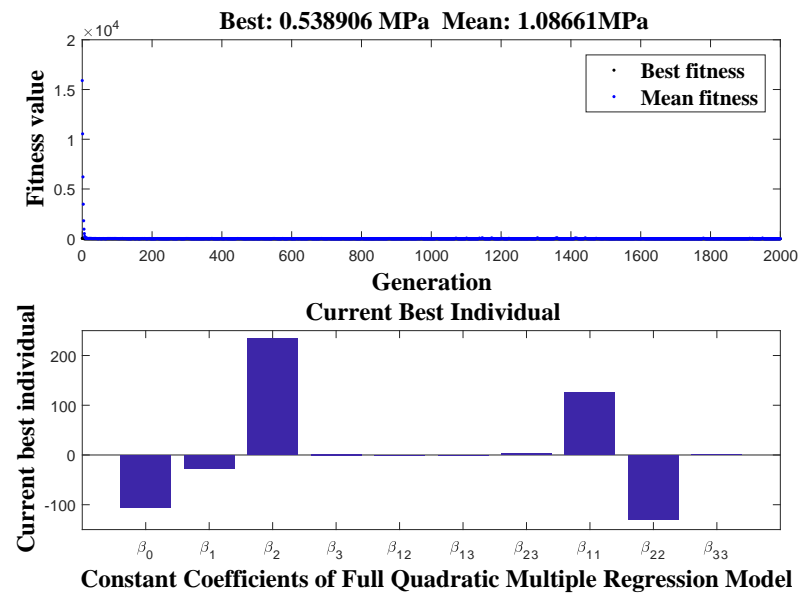


Fig. 7. Performance of the GA in determining the ten constant coefficients for the full quadratic regression model involving three input variables. The upper diagram shows the fitness values over multiple GA iterations (RMSE).

Fig. 7 displays the performance of GA in determining the constant coefficients of the full quadratic multiple regression model for the range 3 Equiflow AICD. The determined constants have an RMSE value equal to 0.53891 MPa. The lower diagram (Fig. 7) illustrates the determined optimum values of the constants.

The lowest RMSE values achieved for the ranges 2, 3 and range 4 are 0.204, 0.53891 and 0.272 MPa, respectively. The tuned (optimum) coefficients for the full quadratic regression model for ranges 2, 3 and 4 of Eqs. (13), (14) and (15),

respectively:

$$\begin{aligned} \delta P = & 657.86 - 36735\mu - 1318.4\rho - 0.73222q \\ & + 36024\mu\rho + 3.8248\mu\rho + 0.92181\mu\rho \\ & + 5.3452632 * 10^5\mu^2 + 659.81\rho^2 \\ & + 0.00093912q^2 \end{aligned} \quad \text{Range(2)} \tag{13}$$

$$\begin{aligned} \delta P = & -105.824 - 27.220\mu + 233.954\rho - 2.0465q \\ & - 0.752\mu\rho - 1.046\mu\rho + 2.727\mu\rho \\ & + 124.982\mu^2 - 129.689\rho^2 \\ & + 0.00138q^2 \end{aligned} \quad \text{Range(3)} \tag{14}$$

Table 6. Layers of an ANFIS model (Tang, 1993; Ebrahim et al., 2019; Chen et al., 2019; Soroush et al., 2019).

Layer	Function	Formulation
Fuzzification	Converts crisp value into fuzzy value	$O_{1,i} = \gamma_{A_i}(Viscosity)$
		$O_{1,i} = \gamma_{B_i}(Density)$
		$O_{1,i} = \gamma_{C_i}(flowrate)$
		$O_{1,i} = \gamma_{D_i}(Range)$
Product layer	Calculates firing strength (w_i)	$O_{2,i} = w_i = \gamma_{A_i}(Viscosity)\gamma_{B_i}(Density)\gamma_{C_i}(flowrate)\gamma_{D_i}(Range)$
Normalization	Calculate the ratio of firing strength	$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1+w_2+w_3+w_4}$
De-Fuzzification	Calculates the node function	$O_{4,i} = \bar{w}_i DP_i = \bar{w}_i(p_i(Viscosity) + q_i(Density) + s_i(Flowrate) + t_i(Range) + r_i)$
Output	Calculates the overall output	$O_{5,i} = \sum_i \bar{w}_i DP_i$

where γ is membership function (MF), p, q, s, t and r are referred to as consequent parameters. The ANFIS model holds a five-layer process to calculate the overall output and A_i and B_i are fuzzy sets that represents the linguistic labels.

$$\begin{aligned} \delta P = & 3.962 + 96.626\mu + 0.000001\rho - 4.094q \\ & - 112.595\mu\rho - 0.3006\mu\rho + 4.8506\mu\rho \\ & + 6.978\mu^2 - 5.383\rho^2 \\ & + 0.00785q^2 \qquad \qquad \qquad Range(4) \end{aligned} \tag{15}$$

5. Adaptive network fuzzy inference system (ANFIS)

ANFIS is a machine learning technique first developed by Tang (1993). It is a hybrid learning procedure which integrates both human knowledge (fuzzy logic, if-then rules) and artificial neural network principles. The *if – then* rules are used to construct a logical relationship between inputs and output of the model (Sugeno and Kang, 1988; Mathur et al., 2016). Table 6 contains the layers, their functions and their outputs for a TSK DP model involving two inputs, viscosity and density, and one output, DP (Jang, 1993).

5.1 ANFIS model development

The data set evaluated by the mathematical and regression models is used to establish a four-input-variable ANFIS model for the Equiflow AICDs. The data records are allocated 80% to the training subset and 20% to the independent testing subset. The Takagi, Sugeno and Kang (TSK) fuzzy inference system method is followed in building the ANFIS model. A hybrid-learning algorithm is employed to train the ANFIS’s adaptive component (Habibi et al., 2018; Ehsanollah et al., 2020). As is the case for the other models, range of AICD, viscosity, density and flow rate are the four input variables and differential pressure is the dependent variable. The structure of the developed ANFIS model is summarized in Table 7.

The RMSE of the developed ANFIS model is equal to 0.18993 MPa for the training subset. Other membership functions were tested for possible use to develop ANFIS models. Table 8 displays the RMSE of the developed ANFIS models for the training and testing data subsets and their ability to identify the optimum viscosity accurately. The **trapmf**, **gaussmf** and **pimf** membership functions are the only membership functions able to identify optimum viscosity accurately.

Table 7. Summary of the four-input-variable ANFIS model developed for AICDs evaluation.

ANFIS components	Value/Type
Type	‘Sugeno’
Version	2
	Range 1
	Viscosity 2
	Density 2
	Flow rate 2
Number of membership functions	
Number of input variables	4
Number of output variables	1
Number of fuzzy rules	8
Number of non-linear parameters	14
Number of linear parameters	40
Type of membership function	“gaussmf”
“And” method applied	“prod”
“Or” method applied	“probor”
Implication method applied	“prod”
Aggregation method applied	“sum”
De-fuzzification method	“wtaver” (weighted average)

Table 8. RMSE values of ANFIS models using various membership functions.

No.	Membership function	Train	Test	Optimum viscosity
1	trimf	0.81	0.57	×
2	trapmf	0.47	0.51	✓
3	gbellmf	0.21	0.24	×
4	gaussmf	0.18	0.20	✓
5	gauss2mf	0.28	0.31	×
6	pimf	0.51	0.49	✓
7	dsigmf	0.50	0.48	×
8	psigmf	0.50	0.48	×

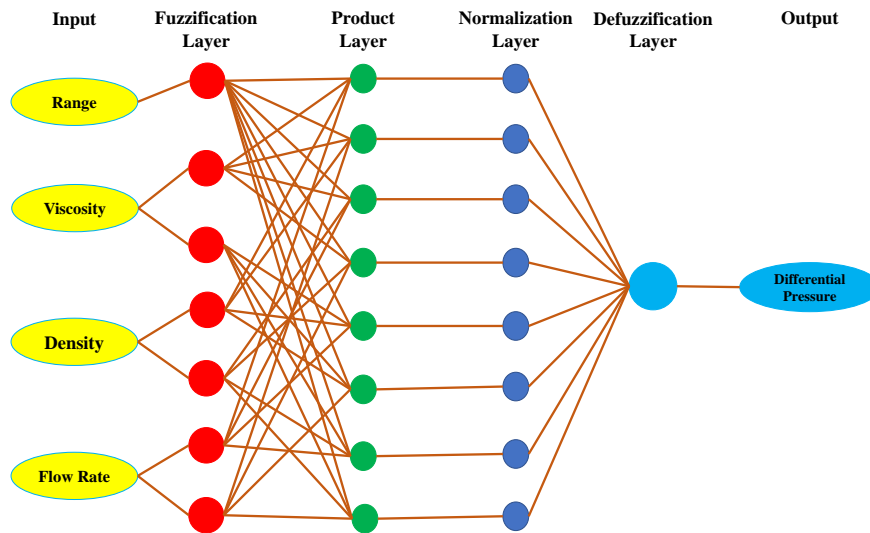


Fig. 8. Structure of the four-input-variable ANFIS model developed for AICDs evaluation.

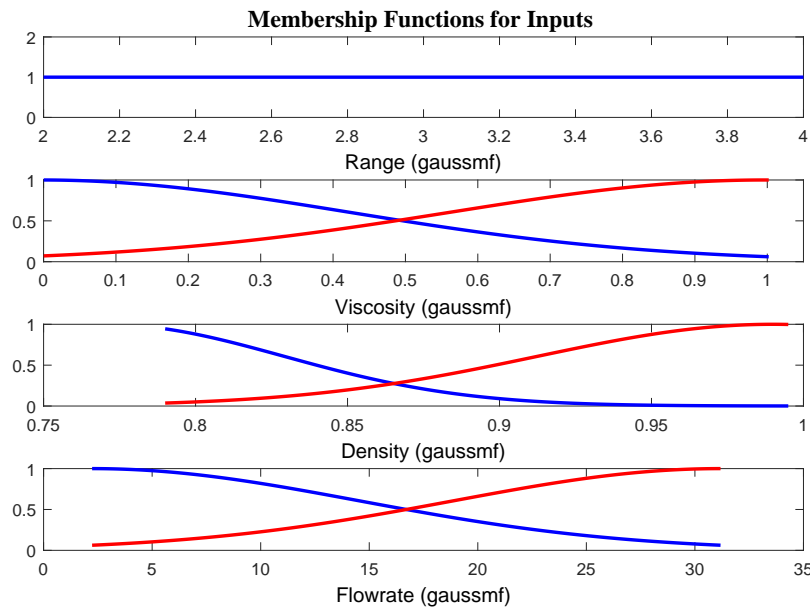


Fig. 9. Membership functions of developed four-input-variable ANFIS model using the hybrid-learning algorithm.

tely. As the RMSE value achieved by **gaussmf** for the testing data is less than **trapmf** and **pimf**, the **gaussmf** membership function was preferred for this analysis.

The membership functions are of the Gaussian type, referred to here as the *gaussmf*, involving 2 constants, *c* and σ , defined in Eq. (16):

$$f(x; \sigma, c) = e^{-\frac{(x-c)^2}{2\sigma^2}} \quad (16)$$

The linguistic expressions for the input data record variable values are binary: either low (L) or high (H). Such simplistic linguistic descriptions can though combine to express the relationships between input variables and the dependent variable in terms of a set of fuzzy IF-THEN rules. The IF-Then rules are formulated using the variable values associated with the

data records included in the training subset. A sample of these rules is as follow:

if Range is (3) Viscosity is (L) and Density is (H) and FlowRate is (H) then...

Differential Pressure = f (Range, Viscosity, Density and FlowRate) is (H)

Fig. 8 illustrates diagrammatically the structure of the four-input-variable ANFIS model developed for AICDs evaluation.

The linguistic labels and corresponding membership functions of the developed ANFIS model are summarized in Table 9. The RMSE of the developed ANFIS model is equal to 0.18993 MPa for the training subset.

Fig. 9 shows the membership functions of input variables in the developed four-input-variable ANFIS model.

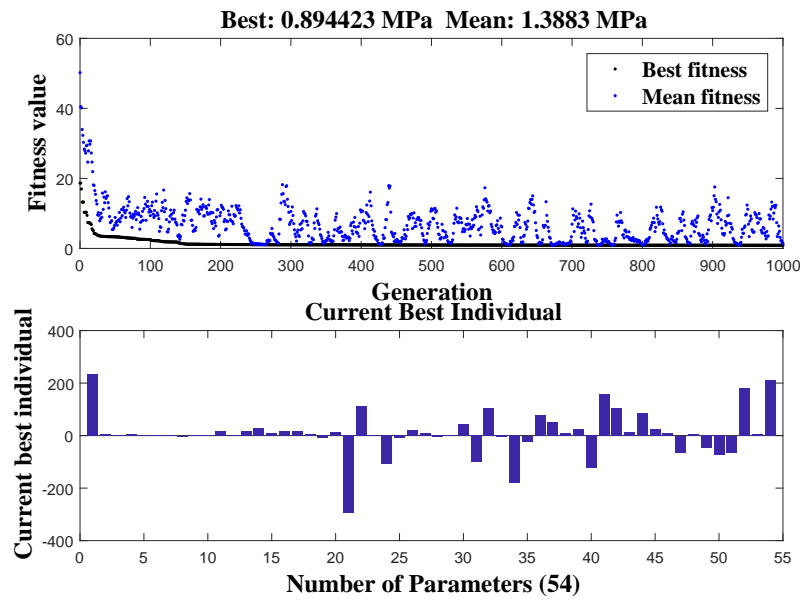


Fig. 10. GA performance in determining the optimum coefficient values for the hybrid four-variable-input ANFIS-GA model. The upper diagram shows the fitness values over multiple GA iterations (RMSE).

Table 9. ANFIS model linguistic labels and corresponding membership functions.

Parameter	Linguistic term	Parameters of membership function	
		σ	c
Range	Range number	inf	4
Viscosity	Low	0.4242	-0.0028
	high	0.4332	0.9961
Density	Low	0.0593	0.7699
	high	0.0776	0.9899
Flow rate	Low	12.2914	2.2347
	high	12.2908	31.1803

5.2 ANFIS-GA model

The developed ANFIS model, described in the previous section, is hybridized with the GA optimizer in an attempt improve the fuzzy inference system coefficients of the ANFIS. Hybridized ANFIS-GA models have been applied with some success in various fields (Kumar et al., 2019; Deshwal et al., 2020; Kumar and Hynes, 2020). Linear and non-linear coefficients of the FIS are key model control parameters responsive to optimization. Table 7 indicates that the developed four-input-variable ANFIS model involves 14 non-linear coefficients and 40 linear coefficients. GA was employed to optimize all 54 of these coefficients associated with the developed ANFIS model. The lower-limit and upper-limit boundary constraints for of these coefficients were set to -1000 and 1000, respectively. The GA control parameters and function applied are those listed in Table 4 and the maximum number of iterations is set to 1000.

The first step in evaluating the ANFIS-GA hybrid model

is to generate an initial population of feasible solutions with a range of FIS coefficients within the constraint boundaries. Each member of the initial population represents a set of membership functions coefficients. These coefficients are used to compute the dependent variable predictions from the ANFIS model. A fitness function value is then calculated for each set of membership functions in terms of the RMSE objective function as expressed in Eq. (4). Based on their RMSE values, the population members are ranked and the GA goes on to apply its crossover, mutation and elitism adaptations to construct a new population for the next iteration. Through multiple successive iterations the GA model converges to an optimum solution or is halted after 1000 iterations and the best solution (lowest RMSE achieved) is selected to provide the optimum FIS coefficient values. The optimum coefficient values determined achieve an RMSE value equal to 0.894423 MPa.

Fig. 10 displays the performance of the GA in determining the optimum coefficients for the ANFIS model using the training subset of data records. The optimum coefficient values determined achieve an RMSE value equal to 0.894423 MPa. The lower diagram in Fig. 10 graphically displays the optimized values of the 54 ANFIS coefficients.

Table 10 summarizes the optimum linguistic labels and their associated MFs.

Fig. 11 shows the membership functions for the optimized hybrid three-input-variable ANFIS-GA model.

The results show that ANFIS's own hybrid algorithm performs better than the GA optimizer in determining the optimized values of the 54 coefficients of ANFIS model. As it is seen, the RMSE value of the designed ANFIS model for training data is equal to 0.189 MPa while, the RMSE value of ANFIS-GA model is equal to 0.894 MPa. Moreover, the ANFIS-GA is time consuming to implement. This indicates that the ANFIS-GA hybrid model performs less well than the

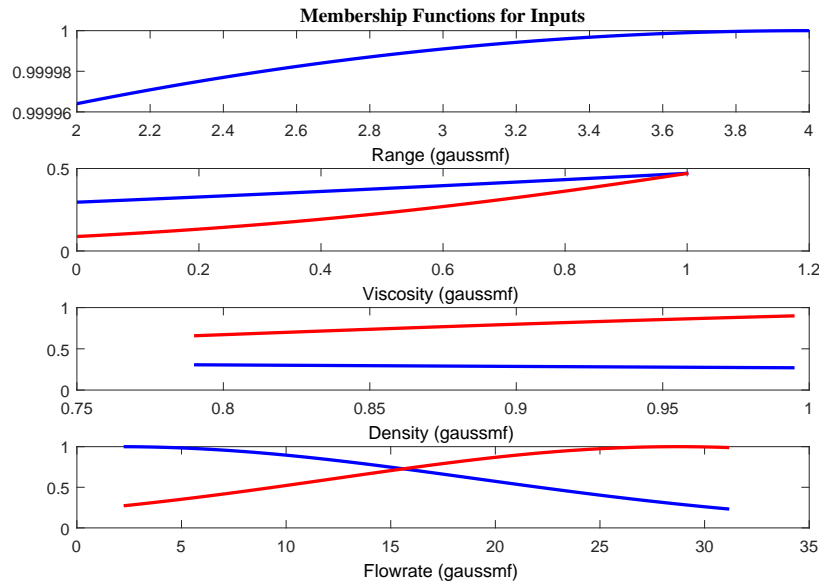


Fig. 11. Membership functions of developed four-input-variable ANFIS-GA model using the genetic algorithm.

Table 10. Hybrid ANFIS-GA linguistic labels and associated membership-function values.

Parameter	Linguistic term	Parameters of membership function	
		σ	c
Range	Range number	235.825	4
Viscosity	Low	3.0016	4.6854
	high	1.0224	2.2559
Density	Low	2.5791	-3.1773
	high	-0.4509	1.2024
Flow rate	Low	17.0902	1.9655
	high	16.4178	28.7138

straightforward ANFIS model for the AICDs data set evaluated.

6. Artificial neural network (ANN)

ANNs are simplified mathematical models mimic in a simplistic way the neural network functionality of animal brains. ANN's are populated by numerous elementary processing units, referred to as neurons, arranged in layers, some of them conducting hidden processing calculations. In most ANN configurations the neurons are assembled in parallel layers passing information from left to right through the network (Mehrotra et al., 1997). At each PE the weighted input signals are summed and the layer bias value is added. I_j refers to the combined input allocated to each hidden layer neuron. f is a linear or non-linear transfer or activation function and y_i describes the transfer-function-adjusted output from one PE fed forward to a neuron in the next layer. Eqs. (17) and (18) summarize formulaically the data adjustments made (Shahin et al., 2001):

$$I_j = \sum_{i=1}^n W_{ji}X_i + \theta_j \quad \text{Summation of combined inputs} \quad (17)$$

$$y_i = f(I_j) \quad \text{Transfer-function adjustment} \quad (18)$$

Various transfer functions are applied to ANNs in practice and they can from layer to layer. They are selected using sensitivity analysis and depend on the complexity of the data set being modeled. Common non-linear transfer functions applied to ANN are “logsig” and/or “tansig”. Common linear transfer functions applied to ANN are “poslin” and “purelin”. The hyperbolic tangent sigmoid function (tansig) is expressed as Eq. (19) (Javad and Nargas, 2010):

$$f = \frac{1}{1 + e^{-av}} - 1 \quad (19)$$

where a is a constant and v represents the weighted-and-bias-adjusted sum (Eq. (17)) of the inputs to a neuron (processing unit). ANN training involves establishing the optimum values for weights and biases and sensitivity analysis to evaluate the effectiveness of different transfer functions and number of hidden layers on dependent value prediction accuracy.

6.1 ANN methodology applied to AICD

The ANN developed for the AICDs model consists of a three-layered feed-forward with Levenberg-Marquardt (LM) algorithm. Training of the AICD ANN is performed using the LM algorithm instead of the standard back-propagation algorithm. This selection was made because the LM algorithm is known to compute more quickly for moderate-sized training subsets in feed-forward neural networks (Bao et al., 2019; Bui et al., 2019; Sun and Huang, 2020). Its single hidden layer has just eight neurons and hyperbolic tangent sigmoid transfer function which is known as “tansig” is applied to the ANN model as transfer function in input and output layers.

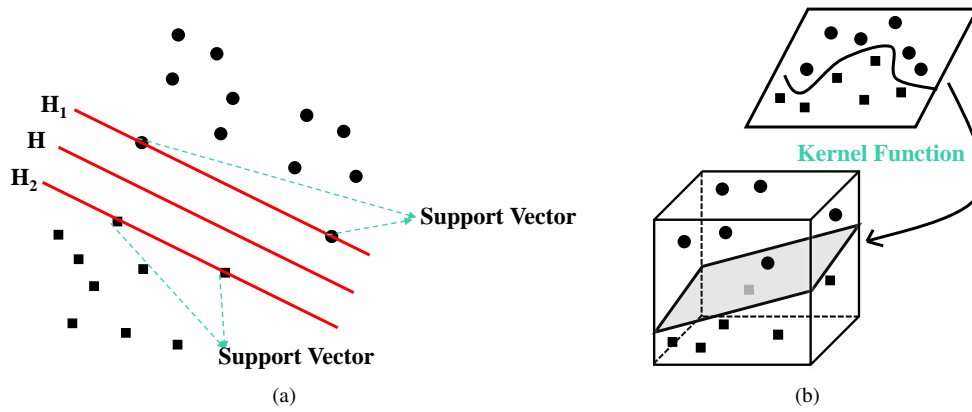


Fig. 12. Schematic diagram of illustrating the fundamental principles of SVMs (a) a classification example between two datasets in 2-D space (b) a kernel function transforming a 3-D data distribution onto a 2D plane that enables the data sets to be more clearly classified (Huang and Zhao, 2018).

As for the other models 80% of the data records are assigned to an ANN training subset and the remaining 20% of data records are held independently a testing subset to verify that the trained model works with data not seen during the training procedure. The AICD ANN developed on the basis described and evaluated for the training subset achieved a prediction error measured for the differential pressure in terms of RMSE equal to 0.17212 MPa.

7. Support-Vector machine (SVM)

SVM is a supervised machine-learning technique developed in 1995 by Vapnik et al. based on a solution for the separable bipartition problem established originally by AT and T Bell laboratories (Vapnik, 1995). It is a method that is applicable for solving either regression or classification problems using a structural-risk minimization approach (Al Amrani et al., 2018; Pham et al., 2018).

SVMs employ kernel functions to establish an optimal classification hyperplane. The square and point symbols plotted in Fig. 12(a) show two groups of data where H represents the classification line that distinguishes the two groups of data, and H1 and H2 (support vectors) are lines parallel to H. The support vectors are drawn close to the data points from each set in vicinity of H classification line and the space between them is referred to as the classification margin. The hyperplane target is the optimum classification solution, such that the two groups of data points are separated from each other and the classification margin is maximized. Fig. 12(b) shows diagrammatically how the kernel function transforms the input data into higher dimensional space eventually facilitating linear classification between the data groups (Huang and Zhao, 2018).

The training subset of data records can be defined in terms of x_i, y_i & ($i = 1, 2, 3, \dots, n$) where n refers to the number of data records in the subset, x_i refers to the input variable values for the i^{th} data record, and y_i refers to the dependent variable values for the i^{th} data record. The SVM optimization function expressed by Eqs. (20) and (21) must be solved in order to locate the optimal hyperplane for classification purposes

(Colkesen et al., 2016; Pham et al., 2018):

$$\text{Minimize : } \frac{1}{2} \|\omega\|^2 + E \sum_{i=1}^n v_i + v_i^* \quad (20)$$

$$\text{Subject to : } \begin{cases} y_i - \omega\varphi(x_i) - u \leq a + v_i \\ \omega\varphi(x_i) + u - y_i \leq a + v_i^* \\ v_i, v_i^* \geq 0 \end{cases} \quad (21)$$

where E is defined as penalty parameter, ω is a weight vector, $\varphi(x_i)$ is kernel function, α is defined as the precision SVM-control parameter, u is bias term and v is defined as the slack variable. Trial and error sensitivity analysis is required to select the most appropriate kernel function to optimize the SVM solution (Burges, 1998). The Gaussian kernel function defined by Eq. (22) was selected on that basis for evaluating the AICD data set (Colkesen et al., 2016; Pham et al., 2018).

$$\exp\left(\frac{-\|x_i - y_i\|^2}{2\sigma^2}\right) \quad (22)$$

The classification decision function of the SVM algorithm is then expressed by Eq. (23).

$$f(x_i) = \text{sign} \left[\sum_{i=1}^n \eta_i \gamma_i \exp\left(\frac{-\|x_i - y_i\|^2}{2\sigma^2}\right) \right] \quad (23)$$

where η_i is a Lagrange multiplier and γ is kernel function (Pham et al., 2018). The SVM model optimizer applied to the AICD data set is “gridsearch” that randomly searches the feasible solution space, using uniform sampling without replacement from a grid. In order to escape from local minima, the *expected improvement – plus* is used as an acquisition function constraint. The algorithm is executed for 2000 iterations. The minimum RMSE value achieved by the developed SVM model is equal to 0.12189 MPa for the training subset. Fig. 13 displays the convergence of SVM model achieved for the AICD training subset.

8. Results: Comparison of AICD differential prediction accuracy

The optimum trained solutions for each of the mathematical, regression and machine learning models applied to the

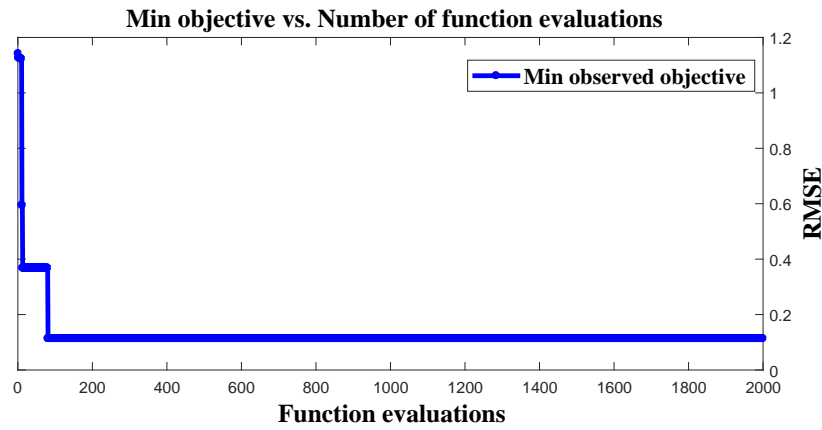


Fig. 13. Minimum objective function value versus number of iterations for developed SVM model.

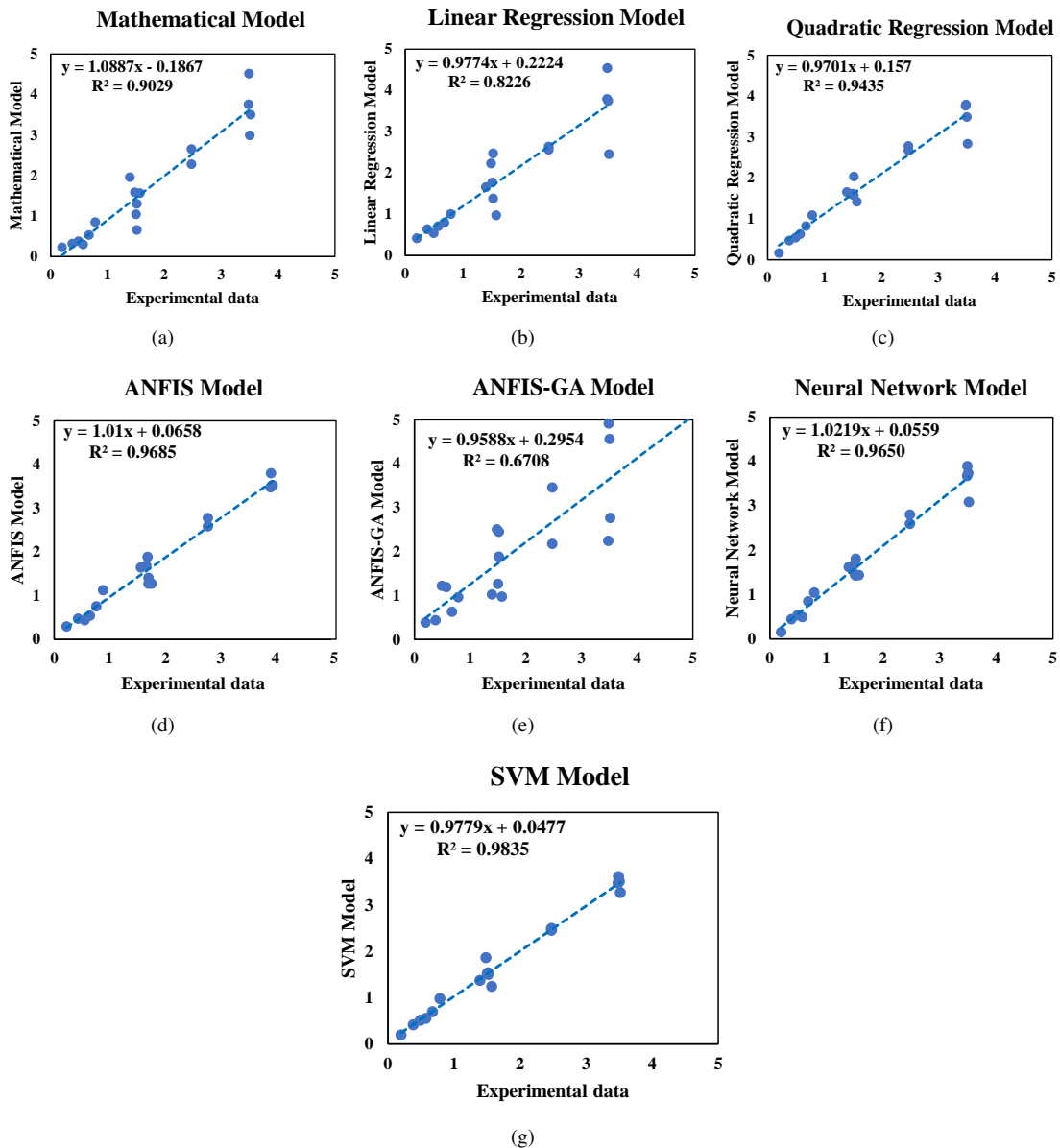


Fig. 14. Predicted versus experimental differential pressure (MPa) for (a) Mathematical, (b) MLR, (c) Full Quadratic Multiple Regression, (d) ANFIS, (e) ANFIS-GA, (f) ANN, (g) SVM models. Linear regression lines and coefficients of determination between predicted and experimental data are displayed.

AICD testing subset are compared graphically in Fig. 14 in terms of predicted versus experimental differential pressure (dependent variable) values. Linear regression lines and coefficient of determination (R^2) are displayed in Fig. 14 to highlight the statistical confidence in the predictions made by each model. In terms of RMSE and R^2 values (Fig. 14), the SVM model ($R^2 = 0.98$) clearly provides the most accurate AICD differential-pressure predictions for the testing subset, whereas the linear regression model ($R^2 = 0.82$) provides the least accurate differential-pressure predictions among mathematical models and ANFIS-GA model ($R^2 = 0.67$) also provides the least performance among machine learning models. For the SVM model, ANFIS model and ANN model the testing subset data record predictions (versus experimental data) lie close to a straight line, which, if extended passes close to the origin. This provides confidence in these models and suggests that the models are not likely to be overfitting the data set. The trained ANFIS model ($R^2 = 0.968$) displays better prediction performance than the trained ANFIS-GA model ($R^2 = 0.67$) when applied to the testing subset (Fig. 14).

The AICD differential-pressure prediction accuracy and reliability of the developed model is further assessed by comparing four widely used statistical prediction accuracy metrics for the training- and testing-subset predictions. These prediction-accuracy metric values compared are root mean squared error (RMSE; Eq. (24)), coefficient of determination (R^2 ; Eq. (25)), average absolute relative deviation (AARD; Eq. (26)) and median absolute deviation (MAD; Eq. (27)).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\delta P_i^e - \delta P_i^p)^2} \quad (24)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (\delta P_i^p - \delta P_i^e)^2}{\sum_{i=1}^N (\delta P_i^p - \bar{\delta P}^e)^2} \quad (25)$$

$$AARD = \frac{100}{N} \sum_{i=1}^N \left| \frac{\delta P_i^e - \delta P_i^p}{\delta P_i^e} \right| \quad (26)$$

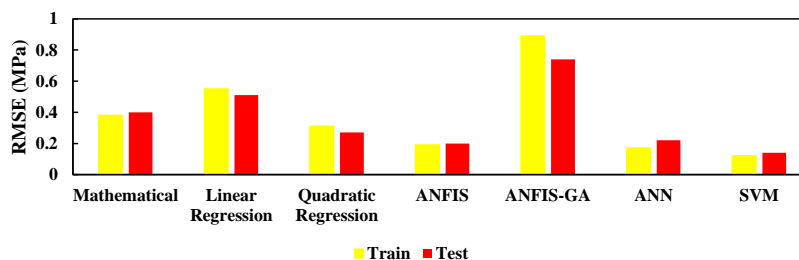
$$MAD = \frac{1}{N} \sum_{i=1}^N |\delta P_i^p - \delta P^e| \quad (27)$$

The values of these four statistical prediction-accuracy metrics for developed models are provided in Table 11. The best prediction accuracy achieved by developed models can be distinctive based on the high values of R^2 and low values of RMSE, AARD and MAD. The SVM model demonstrates the best performance in terms of all four of these metrics.

Fig. 15 depicts the graphical comparison of developed models in terms of RMSE, R^2 , AARD and MAD respectively. Since the SVM has the lowest values of RMSE, AARD and MAD and highest value of R^2 for testing data, it is the most accurate model among other models in estimation of differential pressure of Equiflow AICD. The second most accurate model is ANN model (Total RMSE = 0.1825) and the third most accurate model is ANFIS model (Total RMSE = 0.1931) that shows slightly lower precision than ANN model for total data. The lowest performance among developed models refers to ANFIS-GA model. As mentioned, the ANFIS-GA model takes longer to develop and compute than the other models considered, especially ANFIS model, on the other hand statistical analyses represent that ANFIS-GA exhibits lower prediction performance than ANFIS model. It can be concluded that using GA to adjust FIS parameters of

Table 11. RMSE, R^2 , AARD and MAD for AICD differential-pressure prediction accuracy achieved by the developed models.

Prediction model	RMSE (MPa)		R^2		AARD (%)		MAD (MPa)	
	Train	Test	Train	Test	Train	Test	Train	Test
Mathematical	0.38	0.40	0.94	0.90	24.13	19.28	0.26	0.28
MLR	0.55	0.51	0.84	0.82	63.91	30.02	0.40	0.38
Full quadratic	0.31	0.27	0.95	0.94	31.05	14.27	0.21	0.20
ANFIS	0.19	0.20	0.98	0.97	18.28	14.97	0.12	0.16
ANFIS-GA	0.89	0.74	0.70	0.67	91.27	45.05	0.68	0.62
ANN	0.17	0.22	0.98	0.96	9.82	13.31	0.10	0.18
SVM	0.12	0.14	0.99	0.98	6.768	6.16	0.06	0.08



(a)

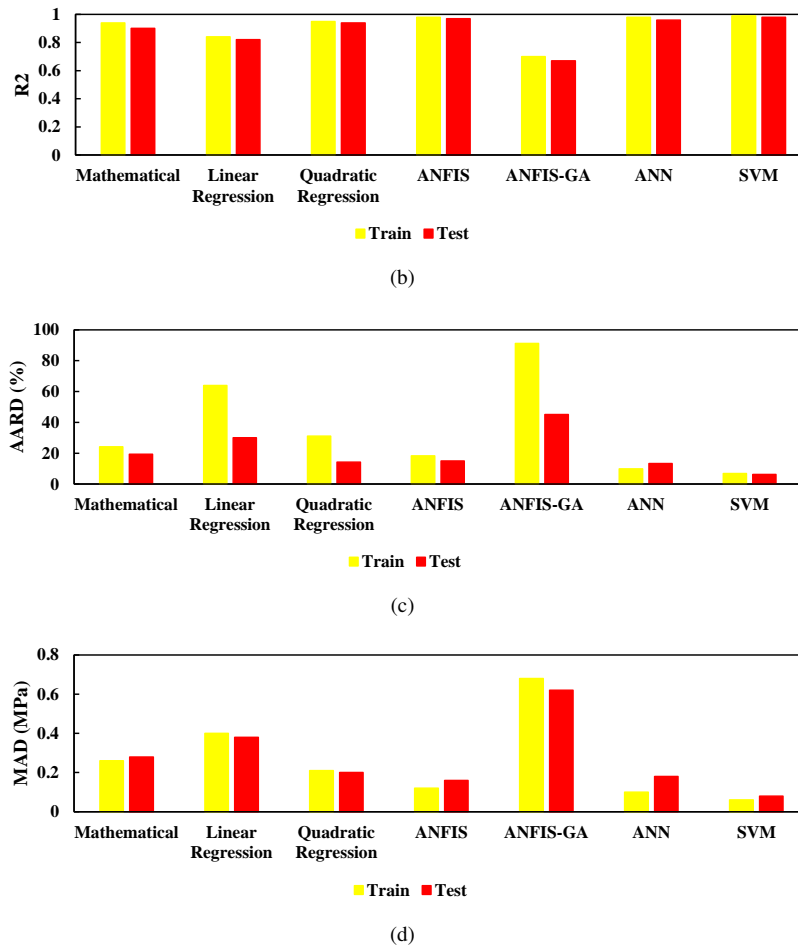


Fig. 15. Graphical comparison of (a) RMSE values (b) R^2 values (c) AARD values (d) MAD values for developed models.

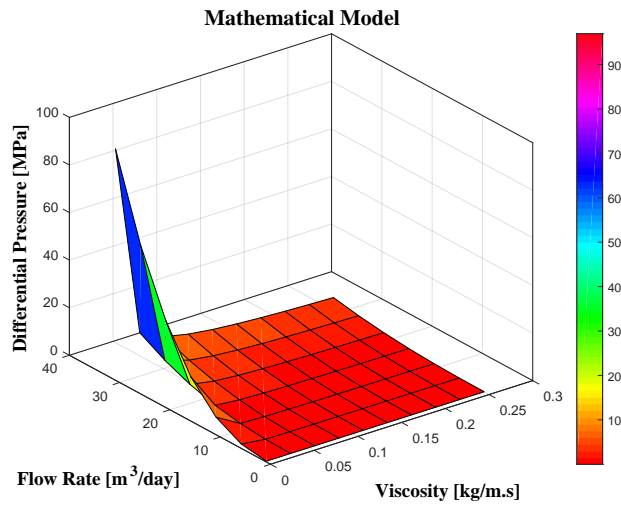
ANFIS model failed in enhancing performance of this type of ANFIS model and hybrid algorithm performs better than GA to determine the ANFIS parameters.

9. Discussion: Consideration of AICD differential prediction reliability

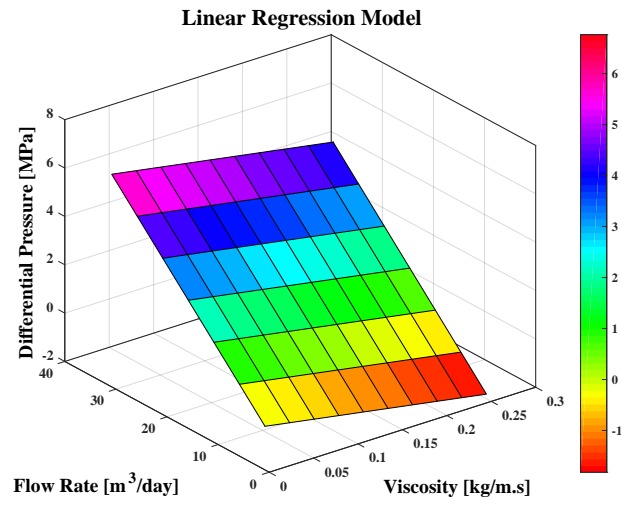
The Equiflow AICD is available in four different designs (Range 1 to Range 4). Each design is configured to work effectively in specific viscosity ranges spanning the spectrum of light to heavy crude oil typically found in subsurface reservoirs. This means that for each design and at a specified differential-pressure value, the flow rate gradually increases as viscosity increases over about a specific 1-cp value increment until it reaches a maximum flow rate. Beyond that maximum flow rate, at higher viscosities, flow rate for that AICD design will slowly decline. Fig. 3 illustrates such a trend with the experimental data, identifying the device tested as an Equiflow Range 3 AICD, because at a given differential pressure by increasing the oil viscosity from 1 to 99 cp the flow rate increases, and then flow rate decreases when oil viscosity increases to 229 cp. These observed trends (Fig. 3) demonstrate that the specific AICD device tested works optimally with crude oil of viscosity between 1 and 99 cp.

Fig. 16 shows the relationships between viscosity, flowrate

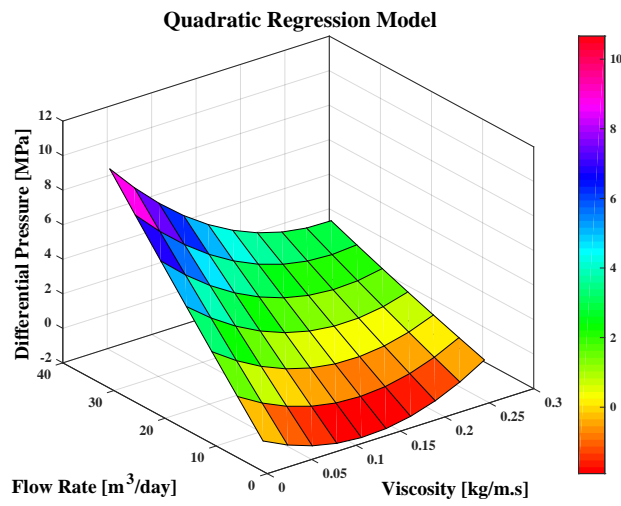
and differential pressure of Equiflow AICD using developed models when crude oil density is equal to 0.87 gr/cm^3 . Fig. 16(a) and Fig. 16(b) show that the AICD performance model and the multi-linear regression models are unable to accurately reflect the optimum performance design concept built into an Equiflow AICD. For any given differential pressure value, as the viscosity increases, the flow rate also increases with these two models, which is incorrect and leads to substantial prediction errors. The machine-learning models are however able to capture the key design behavior of AICDs. A comparison between Fig. 16(d) and Fig. 16(f) shows that both ANN and ANFIS models predict differential pressure accurately and both models are able to display the optimum viscosity of Equiflow AICD. Fig. 16(g) shows that the trained SVM model is able to accurately capture the AICD design concept of optimum viscosity in a smooth manner. The SVM model is able to provide reliably accurate differential-pressure predictions making it superior to the other models developed and tested. Fig. 16(e) shows that ANFIS-GA model is able to show the optimum viscosity but it cannot be used as a reliable model to predict the differential pressure of Equiflow AICD compared to other developed models. According to Fig. 16(c), although it seems that quadratic multiple linear regression model is able to model the optimum viscosity of the



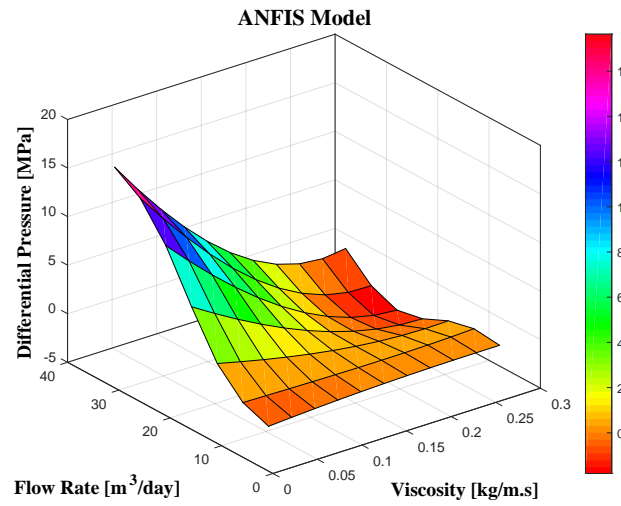
(a)



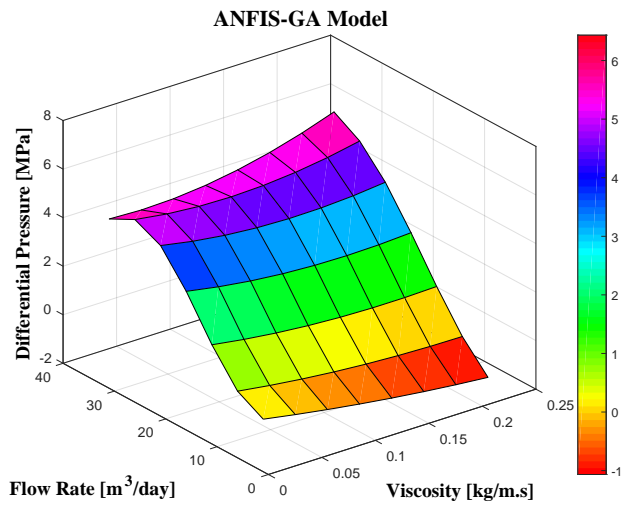
(b)



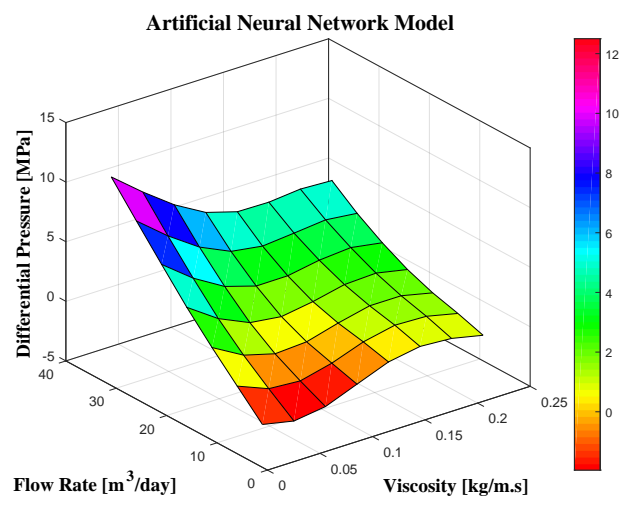
(c)



(d)



(e)



(f)

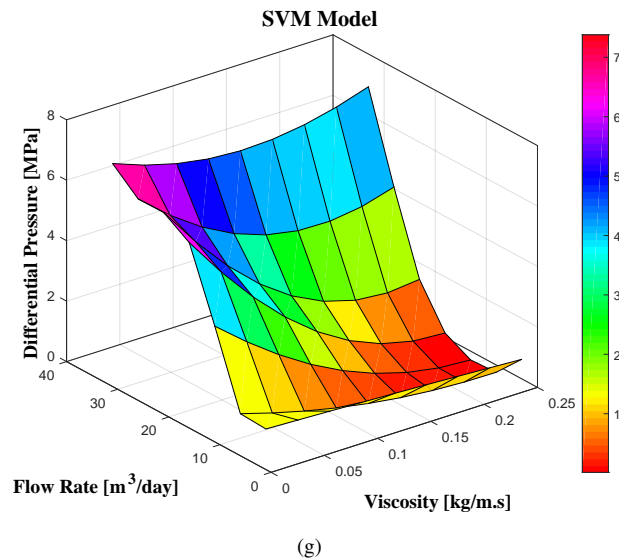


Fig. 16. Relationship between viscosity, flow rate and differential pressure at crude oil density of 0.87 g/cm^3 using (a) the optimized AICD performance model (b) Multiple Linear Regression (c) Full Quadratic Multiple Regression (d) Developed ANFIS (e) Optimized ANFIS-GA (f) developed ANN (g) designed SVM.

Equipflow AICD accurately, the linear structure of this model causes some errors in prediction of differential pressure of the Equipflow AICDs because the relationship between flow rate and differential pressure is not linear.

It is apparent that the machine learning models perform substantially better than mathematical models. However, key factors that affect the accuracy of machine learning models are the number of experimentally derived training and testing data points available and the distribution of those data points. When the available data points are too few or poorly distributed, machine learning models may perform in an inferior way to the mathematical models. In some cases, very few data points may generate statistically unsound results that are not physically justifiable. In such cases where the data set is very small, mathematical models should be relied upon to model the performance of Equipflow AICDs.

10. Conclusions

In this study, several machine-learning models including ANN, ANFIS, ANFIS-GA and SVM, together with multiple regressions and a mathematical model are developed and implemented to predict differential pressure for ranges 2, 3 and 4 Equipflow downhole autonomous inflow control devices (AICDs). A set of ninety experimental data records were compiled in field-like conditions to develop the models. These data are divided into two subsets: a training subset of 80% of the data records were utilized to construct and train the models; and, a testing subset made up of the remaining 20% of data records. The testing subset was utilized to independently test and verify the performance and accuracy of the constructed models. Viscosity, density, flow rate and the range of AICDs are the input variables and the differential pressure is the prediction output (dependent variable) of the models. Statistical and graphical analyses reveal that all the machine-learning models perform better than multiple

regression mathematical models. The full quadratic multiple regression model is the only mathematical model able to deliver credible differential-pressure prediction performance for the Equipflow AICDs. The most accurate and reliable model relative to the other developed models is the SVM model. The ANN model performed better than the ANFIS model. A genetic optimization algorithm (GA) worked well in finding the optimum tuning parameters for the mathematical models but did not improve on the prediction performance of ANFIS when evaluated in a hybrid ANFIS-GA configuration.

Acknowledgement

The authors thank their respective institutions for supporting this research.

Conflicts of interest

The authors declare no competing interest.

Open Access This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- Aakre, H., Halvorsen, B., Werswick, B., et al. Smart well with autonomous inflow control valve technology. Paper SPE 164348 Presented at Middle East Oil and Gas Show and Conference, Manama, Bahrain, 10-13 March, 2013.
- Aakre, H., Halvorsen, B., Werswick, B., et al. Autonomous inflow control valve for heavy and extra-heavy oil. Paper SPE 71141 Presented at Heavy and Extra Heavy Oil Conference, Latin America, Medellín, Colombia, 24-26 September, 2014.
- Abad, A. R. B., Mousavi, S., Mohamadian, N., et al. Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate

- reservoirs. *Journal of Natural Gas Science and Engineering*, 2021, 95: 104210.
- Aiken, L. S., West, S. G., Pitts, S. C., et al. Multiple linear regression, in *Handbook of Psychology*, edited by Weiner, J. A. Schinka and W. F. Velicer, Second Edition, John Wiley & Sons, Inc, New York, 2012.
- Al Amrani, Y., Lazaar, M., El Kadiri, K. E. Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, 2018, 127: 511-520.
- Al-Khelaiwi, F. T., Davies, D. R. Inflow control devices: Application and value quantification of a developing technology. Paper SPE 108700 Presented at International Oil Conference and Exhibition in Mexico, Veracruz, Mexico, 27-30 June, 2007.
- Andreas, L., Midttveit, O., Gyllensten, A. J., et al. AICD implementation on oseberg H vestflanken 2. Paper SPE 195617 Presented at SPE One Day Seminar, Bergen, Norway, 14 May, 2019.
- Ashrafi, S. B., Anemangely, M., Sabah, M., et al. Application of hybrid artificial neural networks for predicting rate of penetration (ROP): A case study from Marun oil field. *Journal of Petroleum Science and Engineering*, 2019, 175: 604-623.
- Bao, H., Wang, J., Li, J., et al. Effects of corn straw on dissipation of polycyclic aromatic hydrocarbons and potential application of backpropagation artificial neural network prediction model for PAHs bioremediation. *Ecotoxicology and Environmental Safety*, 2019, 186: 109745.
- Barbosa, L. F. F., Nascimento, A., Mathias, M. H., et al. Machine learning methods applied to drilling rate of penetration prediction and optimization-A review. *Journal of Petroleum Science and Engineering*, 2019, 183: 106332.
- Bui, X-N., Muazu, M. A., Nguyen, H. Optimizing Levenberg-Marquardt backpropagation technique in predicting factor of safety of slopes after two-dimensional OptumG2 analysis. *Engineering with Computers*, 2019, 36: 941-952.
- Burges, C. J. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 1998, 2(2): 121-167.
- Chen, W., Panahi, M., Khosravi, K., et al. Spatial prediction of groundwater potentiality using ANFIS ensembled with teaching-learning-based and biogeography-based optimization. *Journal of Hydrology*, 2019, 572: 435-448.
- Cheng, L., Wang, D., Cao, R., et al. The influence of hydraulic fractures on oil recovery by water flooding processes in tight oil reservoirs: an experimental and numerical approach. *Journal of Petroleum Science and Engineering*, 2020, 185: 106572.
- Chochua, G., Rudic, A., Kumar, A., et al. Cyclone type autonomous inflow control device for water and gas control: Simulation-driven design. Paper SPE 192723 Presented at Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 12-15 November, 2018.
- Ciulla, G., D'Amico, A. Building energy performance forecasting: A multiple linear regression approach. *Applied Energy*, 2019, 253: 113500.
- Colkesen, I., Sahin, E. K., Kavzoglu, T. Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression. *Journal of African Earth Sciences*, 2016, 118: 53-64.
- Deshwal, S., Kumar, A., Chhabra, D. Exercising hybrid statistical tools GA-RSM, GA-ANN and GA-ANFIS to optimize FDM process parameters for tensile strength improvement. *CIRP Journal of Manufacturing Science and Technology*, 2020, 31: 189-199.
- Ebrahim, S., Mesbah, M., Hajilari, N., et al. ANFIS modeling for prediction of CO₂ solubility in potassium and sodium based amino acid salt solutions. *Journal of Environmental Chemical Engineering*, 2019, 7(1): 102925.
- Ehsanollah, H., Salehi, M., Yadegarfar, G., et al. Optimization of the ANFIS using a genetic algorithm for physical work rate classification. *International Journal of Occupational Safety and Ergonomics*, 2020, 26(3): 436-443.
- Elkatatny S. Real-time prediction of rate of penetration while drilling complex lithologies using artificial intelligence techniques. *Ain Shams Engineering Journal*, 2021, 12(1): 917-926.
- Eltaher, E. M. K. Modelling and applications of autonomous flow control devices. Edinburgh, Heriot-Watt University, 2017.
- Eltaher, E. Muradov, K., Davies, D., et al. Autonomous flow control device modelling and completion optimization. *Journal of Petroleum Science and Engineering*, 2019, 17: 995-1009.
- Fripp, M., Zhao, L., Least, B. The theory of a fluidic diode autonomous inflow control device. Paper SPE 167415 Presented at SPE Middle East Intelligent Energy Conference and Exhibition, 28-30 October, 2013.
- Gamal, H., Elkatatny, S., Abdurraheem, A. Rock drillability intelligent prediction for a complex lithology using artificial neural network. Paper SPE 202767 Presented at Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 9-12 November, 2020.
- Gao, C. H., Rajeswaran, R. T., Nakagawa, E. Y. A literature review on smart well technology. Paper SPE 106011 Presented at Production and Operations Symposium, Oklahoma City, Oklahoma, USA, 31 March-3 April, 2007.
- Ghorbani, H., Wood, D. A., Choubineh, A., et al. Performance comparison of bubble point pressure from oil PVT data: Several neurocomputing techniques compared. *Experimental and Computational Multiphase Flow*, 2020, 2: 225-246.
- Gimre, J. Efficiency of ICV/ICD systems. Stavanger, University of Stavanger, 2012.
- Glandt, C. A. Reservoir management employing smart wells: A review. *SPE Drilling & Completion*, 2005, 20(4): 281-288.
- Goldberg, D. E., Holland, J. H. Genetic algorithms and machine learning. *Machine Learning*, 1988, 3: 95-99.

- Gurses, S., Chochua, G., Rudic, A., et al. Dynamic modeling and design optimization of cyclonic autonomous inflow control devices. Paper SPE 193824 Presented at SPE Reservoir Simulation Conference, Galverston, Texas, USA, 10-11 April, 2019.
- Habibi, E., Salehi, M., Yadegarfar, G., et al. Optimization of the ANFIS using a genetic algorithm for physical work rate classification. *International Journal of Occupational Safety and Ergonomics*, 2020, 26(3): 436-443.
- Halvorsen, M., Elseth, G., Nævdal, O. M. Increased oil production at Troll by autonomous inflow control with RCP valves. Paper SPE 159634 Presented at SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, 8-10 October, 2012.
- Hazbeh, O., Aghdam, S. K. Y., Ghorbani, H., et al. Comparison of accuracy and computational performance between the machine learning algorithms for rate of penetration in directional drilling well, *Petroleum Research*, 2021, 6(3): 271-282.
- Holland, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge, Massachusetts, USA, The MIT Press, 1992b.
- Holland, J. H. Genetic algorithms. *Scientific American*, 1992a, 267(1): 44-50.
- Huang, Y., Zhao, L. Review on landslide susceptibility mapping using support vector machines. *Catena*, 2018, 165: 520-529.
- Iqbal, F., Iskandar, R., Radwan, E., et al. Autonomous inflow control device-a case study of first successful field trial in GCC for water conformance. Paper SPE 177927 Presented at Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, UAE, 9-12 November, 2015.
- Javad, G., Narges, T. Application of artificial neural networks to the prediction of tunnel boring machine penetration rate. *Mining Science and Technology*, 2010, 20(5): 727-733.
- Jia, C. Breakthrough and significance of unconventional oil and gas to classical petroleum geology theory. *Petroleum Exploration and Development*, 2017, 44(1): 1-10.
- Jovanov, I. *Performance of autonomous inflow control systems*. Stavanger, University of Stavanger, 2016.
- Jovic, S., Anicic, O., Pejovic, B. Management of the wind speed data using adaptive neuro-fuzzy methodology. *Flow Measurement and Instrumentation*, 2016, 50: 201-208.
- Khademi, F., Akbari, M., Jamal, S. M., et al. Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive strength of concrete. *Frontiers of Structural and Civil Engineering*, 2017 11(1): 90-99.
- Kumar, R., Hynes, N. R. J. Prediction and optimization of surface roughness in thermal drilling using integrated ANFIS and GA approach. *Engineering Science and Technology*, 2020, 23(1): 30-41.
- Kumar, V., Kumar, A., Chhabra, D., et al. Improved biobleaching of mixed hardwood pulp and process optimization using novel GA-ANN and GA-ANFIS hybrid statistical tools. *Bioresource technology*, 2019, 271: 274-282.
- Lauritzen, J. E., Martiniussen, I. B. Single and multi-phase flow loop testing results for industry standard inflow control devices. Paper SPE 146347 Presented at SPE Offshore Europe Oil and Gas Conference and Exhibition, Aberdeen, UK, 6-8 September, 2011.
- Least, B., Greci, S., Burkey, R. C., et al. Autonomous ICD single phase testing. Paper SPE 160165 Presented at SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, 8-10 October, 2012.
- Lei, Q., Jackson, M. D., Muggeridge, A. H., et al. Modelling the reservoir-to-tubing pressure drop imposed by multiple autonomous inflow control devices installed in a single completion joint in a horizontal well. *Journal of Petroleum Science and Engineering*, 2020, 189: 106991.
- Mardani-rad, S., Wood, D. A., Zakeri, H. The application of deep learning algorithms to classify subsurface drilling lost circulation severity in large oil field datasets. *SN Applied Sciences*, 2021, 3: 785.
- Mathiesen, V., Werswick, B., Aakre, H., et al. The Autonomous RCP valve - new technology for inflow control in horizontal wells autonomous valve. Paper SPE 145737 Presented at SPE Offshore Europe Oil and Gas Conference and Exhibition, Aberdeen, UK, September 6-8, 2011.
- Mathur, N., Glesk, I., Buis, A. Comparison of adaptive neuro-fuzzy inference system (ANFIS) and Gaussian processes for machine learning (GPML) algorithms for the prediction of skin temperature in lower limb prostheses. *Medical Engineering & Physics*, 2016, 38(10): 1083-1089.
- Mehrotra, K., Mohan, C. K., Ranka, S. *Elements of Artificial Neural Networks*, Cambridge, Massachusetts, USA, The MIT Press, 1997.
- Mehrad, M., Bajolvand, M., Ramezanzadeh, A., et al. Developing a new rigorous drilling rate prediction model using a machine learning technique. *Journal of Petroleum Science and Engineering*, 2020, 192: 107338.
- Mikkelsen, J. K., Norheim, T., Sagatun, S. The Troll story. Paper SPE 171108 Presented at the Offshore Technology Conference, Houston, Texas, 2-5 May, 2005.
- Mitchell, M. *An Introduction to Genetic Algorithms*. Cambridge, Massachusetts, USA, The MIT Press, 1996.
- Moazzeni, A. R., Khomehchi, E. Rain optimization algorithm (ROA): A new metaheuristic method for drilling optimization solutions. *Journal of Petroleum Science and Engineering*, 2020, 195: 107512.
- Mohamadian, N., Ghorbani, H., Wood, D. A., et al. A geomechanical approach to casing collapse prediction in oil and gas wells aided by machine learning. *Journal of Petroleum Science and Engineering*, 2021, 196: 107811.
- Moraveji, M. K., Naderi, M. Drilling rate of penetration prediction and optimization using response surface methodology and bat algorithm. *Journal of Natural Gas Science and Engineering*, 2016, 31: 829-841.
- Nathans, L. L., Oswald, F. L., Nimon, K. *Interpreting multiple linear regression: A guidebook of variable importance*. Practical Assessment, Research, and Evaluation, 2012,

- 17(9): 1-19.
- Ossai, C. I., Duru, U. I. Applications and theoretical perspectives of artificial intelligence in the rate of penetration. *Petroleum*, 2020, in Press, doi: 10.1016/j.petlm.2020.08.004.
- Pedroso, C. A., Latini, C., Araujo, Z., et al. First open-hole gravel pack with acid in ultra deep water, (2020). Paper SPE 199325 Presented at SPE International Conference and Exhibition on Formation Damage Control, Lafayette, Louisiana, USA, 19-21 February, 2020.
- Pham, B. T., Bui, D. T., Prakash, I. Bagging based support vector machines for spatial prediction of landslides. *Environmental Earth Sciences*, 2018, 77(4): 146.
- Porturas, F. Enhanced production with ICD and AICD completions in oil wells: Case studies from Latin America. Paper SPE 181204 Presented at SPE Latin America and Caribbean Heavy and Extra Heavy Oil Conference, Lima, Peru, 19-20 October, 2016.
- Rabiei, A., Sayyad, H., Riazi, M., et al. Determination of dew point pressure in gas condensate reservoirs based on a hybrid neural genetic algorithm. *Fluid Phase Equilibria*, 2015, 387: 38-49.
- Rahul, M., Narinder, S., Yaduvir, S. Genetic Algorithms: Concepts, design for optimization of process controller. *Computer and Information Science*, 2011, 4(2): 39-54.
- Rashid, S., Ghamartale, A., Abbasi, J., et al. Prediction of critical multiphase flow through chokes by using A rigorous artificial neural network method. *Flow Measurement and Instrumentation*, 2019, 69: 101579.
- Sabah, M., Talebkeikhah, M., Wood, D. A., et al. A machine learning approach to predict drilling rate using petrophysical and mud logging data. *Earth Science Informatics*, 2019, 12: 319-339.
- Sadrmomtazi, A., Sobhani, J., Mirgozar, M. A. Modeling compressive strength of EPS lightweight concrete using regression, neural network and ANFIS. *Construction and Building Materials*, 2013, 42: 205-216.
- Sekki, T., Airaksinen, M., Saari, A. Impact of building usage and occupancy on energy consumption in Finnish daycare and school buildings. *Energy and Buildings*, 2015, 105: 247-257.
- Shahin, M. A., Jaksa, M. B., Maier, H. R. Artificial neural network applications in geotechnical engineering. *Australian Geomechanics*, 2020, 36(1): 49-62.
- Somehsaraei, H. N., Hölle, M., Hönen H. A novel approach based on artificial neural network for calibration of multi-hole pressure probes. *Flow Measurement and Instrumentation*, 2020, 73: 101739.
- Soroush, E., Mesbah, M., Hajilary, N., et al. ANFIS modeling for prediction of CO₂ solubility in potassium and sodium based amino acid Salt solutions. *Journal of Environmental Chemical Engineering*, 2019, 7(1): 102925.
- Sugeno, M., Kang, G. T. Structure identification of fuzzy model. *Fuzzy Sets and Systems*, 1988, 28(1): 15-33.
- Sun, W., Huang, C. A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. *Journal of Cleaner Production*, 2020, 243: 118671.
- Tang, J-S. R. ANFIS: adaptive network based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 1993, 23: 515-520.
- Vapnik, V. N. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, USA, 1995.
- Velez-Langs, O. Genetic algorithms in oil industry: An overview. *Journal of petroleum Science and Engineering*, 2005, 47(1-2): 15-22.
- Vickers, N. J. Animal communication: When i'm calling you, will you answer too? *Current Biology*, 2017, 27(14): R713-R715.
- Wood, D. A. A transparent open-box learning network provides insight to complex systems and a performance benchmark for more-opaque machine learning algorithms. *Advances in Geo-Energy Research*, 2018, 2(2): 148-162.
- Yavari, Y., Sabah, M., Khosravanian, R., et al. Application of an adaptive neuro-fuzzy inference system and mathematical rate of penetration models to predicting drilling rate. *Iranian Journal of Oil and Gas Science and Technology*, 2018, 7(3): 73-100.
- Yilmaz, I., Kaynar, O. Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. *Expert Systems with Applications*, 2011, 38(5): 5958-5966.
- Zhang, N., Li, H., Liu, Y., et al. A new autonomous inflow control device designed for a loose sand oil reservoir with bottom water. *Journal of Petroleum Science and Engineering*, 2019, 178: 334-355.
- Zhao, L., Least, B., Greci, S., et al. Fluidic diode autonomous ICD range 2A single-phase testing. Paper SPE 170993 Presented as SPE Oilfield Water Management Conference and Exhibition, Kuwait City, Kuwait, 21-22 April, 2014.
- Zhao, X., Zhou, L., Pu, X., et al. Exploration breakthroughs and geological characteristics of continental shale oil: A case study of the Kongdian Formation in the Cangdong Sag, China. *Marine and Petroleum Geology*, 2019, 102: 544-556.