CrossMark

ORIGINAL ARTICLE

Prediction of non-hydrocarbon gas components in separator by using Hybrid Computational Intelligence models

Tarek Helmy[1] · Muhammad Imtiaz Hossain[1] · Abdulazeez Adbulraheem[2] ·
S. M. Rahman[3] · Md. Rafiul Hassan[1] · Amar Khoukhi[4] · M. Elshafei[4]

Abstract Accurate prediction of non-hydrocarbon (Non-HC) gas components in the gas-oil separators reduces the cost of gas and oil production in petroleum engineering. However, this task is difficult because there is no known relation among the properties of crude oil and the separators. There are studies that attempt to predict hydrocarbons (HCs) components using either Computational Intelligence (CI) techniques or conventional techniques like Equitation-of-State (EOS) and Empirical Correlation (EC). In this paper, we explore the applicability of CI techniques such as Artificial Neural Network, Support Vector Regressions, and Adaptive Neuro-Fuzzy Inference System to predict the Non-HC gas components in gas-oil separator tank. Further, we incorporate Genetic Algorithms (GA) into the Hybrid Computational Intelligence (HCI) models to enhance the accuracy of prediction. GA is used to determine the most favorable values of the tuning parameters in the CI models. The performances of the CI and HCI models are compared with the performance of the conventional techniques like EOS and EC. The experimental results show that accuracy of prediction by CI and HCI models outperform the conventional methods for N_2 and H_2S gas components. Furthermore, the HCI models perform better than the non-optimized CI models while predicting the Non-HC gas components.

Keywords Non-HC gas components prediction · Artificial Neural Network · Support Vector Regression · Adaptive Neuro-Fuzzy System · Hybrid Computational Intelligence · Genetic Algorithms

Abbreviations
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
CC	Correlation coefficient
CI	Computational intelligence
CO_2	Carbon di oxide
EC	Empirical correlations

✉ Tarek Helmy
helmy@kfupm.edu.sa

Muhammad Imtiaz Hossain
mdimtiaz@kfupm.edu.sa

Abdulazeez Adbulraheem
aazeez@kfupm.edu.sa

S. M. Rahman
smrahman@kfupm.edu.sa

Md. Rafiul Hassan
mrhassan@kfupm.edu.sa

Amar Khoukhi
amar@kfupm.edu.sa

M. Elshafei
elshafei@kfupm.edu.sa

[1] Department of Information and Computer Science, King Fahd University of Petroleum and Minerals, Dhahran 31261, KSA

[2] Department of Petroleum Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, KSA

[3] Center for Environment and Water, Research Institute, King Fahd University of Petroleum and Minerals, Dhahran 31261, KSA

[4] Department of Systems Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, KSA

🙋 Springer

EOS	Equation-of-states
FIS	Fuzzy inference system
FL	Fuzzy logic
GA	Genetic algorithm
GOSP	Gas oil separation plant
H_2S	Hydrogen sulfide
HC	Hydrocarbon
HCI	Hybrid computational intelligence
LM	Levenberg–Marquardt
MLP	Multi-layer perceptron
MW	Molecular weight
N_2	Nitrogen
Non-HC	Non-hydrocarbon
P	Pressure (psi)
P_b	Bubble point pressure (psi)
PR-EOS	Peng–Robinson EOS
PEPs	Petroleum engineering problems
RMSE	Root-mean-square error
Rprop	Resilient back-propagation
RT	Reservoir temperature (°F)
SP	Separator pressure (psi)
ST API	Stock Tank American Petroleum Institute
ST	Separator temperature (°F)
Subclust	Subtractive clustering
SVM	Support vector machine

1 Introduction

Hydrocarbon (HC) and non-hydrocarbon (Non-HC) gas components compose crude oil in reservoirs. The oil is extracted from reservoir and then collected in sequential separator tanks in much lower temperature as well as much lower pressure as shown in Fig. 1. Due to huge fall of temperature and pressure, gas components are released from each separator stage of the Gas–Oil Separation Plant (GOSP). The amount of Non-HC gases, e.g., nitrogen (N_2), carbon di oxide (CO_2) and hydrogen sulfide (H_2S) in separators is on average around 5–6 % which is negligible compare to other components. These gases usually vary in a wide range, e.g., N_2 (0–9 %), CO_2 (0–16 %) and H_2S (0–19 %). Furthermore, the amount of mole percent of gas in the output is usually higher in the separator than that of the immediate previous stage, which varies in ranges hence difficult to predict. In the separation process, the mixtures of Non-HC gases are treated as impurities hence not desirable in the remaining stock tank oil [21]. Therefore, Non-HC gas components prediction beforehand will help in cutting down the production cost, maximizing the production efficiency and determining the quality of gas as well as stock tank oil.

1.1 Background and motivation

Usually, the amount of gases and oil are measured in laboratory experiments known as "separator tests". Separator tests involve huge time and expensive equipment hence uneconomical. As an alternative, complicated relations such as EOS and EC are being used in the industry to compute the amount of gases and oils. EOS is useful for description of fluid properties such as pressure–volume–temperature (PVT). Unfortunately there is no single EOS that accurately estimates the properties of all substances under all conditions [48]. Moreover, the EOS has adjustment issues against the phase behavior data of reservoir fluid composition, while the EC has limited accuracy [14]. A number of studies exist that attempted to predict HCs using CI models [8, 14, 15, 45]. However, the performances of the existing CI models are subjective and dependent on selection of optimal parameters. Recently Hossain et al. [25] used only ANN to predict the Non-HC gas components in separator that shows good predictive accuracy and motivates us to explore the applicability of more CI and HCI models. To the best of our knowledge, no other significant studies have been noticed that attempted to predict the Non-HC gases in separator using either CI or HCI models. In this paper we propose HCI models to predict the Non-HC gas components in gas-oil separator and enhance the accuracy of prediction using GA.

1.2 Our contribution

We proposed a GA-based HCI models to predict the amount of Non-HC gases (N_2, CO_2 and H_2S) in a separator at GOSP. It is reported that the prediction accuracy of the existing CI models, e.g., ANN, SVR and ANFIS solely depends on the suitable selection of model's parameters [25]. In this paper, determining the best favorable values of the tuning parameter set of the CI models is achieved through application of GA such that the HCI model becomes more capable of learning the complex relationship between the input and the output parameters accurately compared to the individual CI models without having optimal/sub-optimal parameter sets. Experimental results on the Asian reservoirs samples show that the HCI models outperform of non-optimized CI models while predicting Non-HC gas components in a multi-stage separator of GOSP. Furthermore, the proposed HCI models outperform the conventional methods in predicting at least two Non-HC gas components out of the three, i.e., N_2, CO_2 and H_2S.

2 Literature survey

In recent years several CI models, e.g., ANN, SVR and ANFIS have gained huge popularity in solving various research areas including petroleum engineering. In this

Fig. 1 Oil and gas flow in multi-stage separator
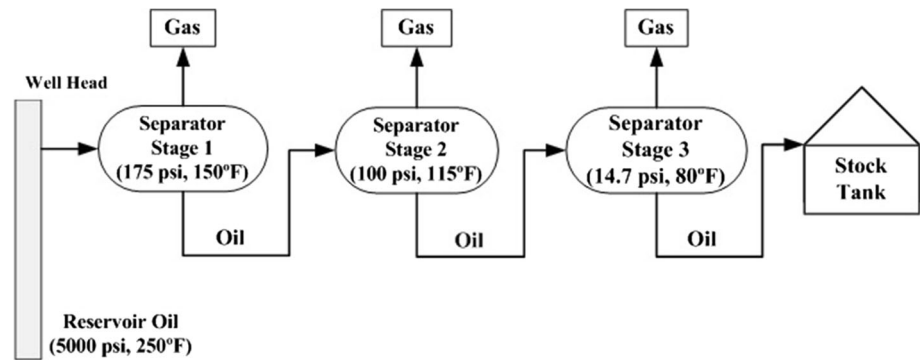
section, we will discuss about the application of CI models to solve petroleum engineering issues including oil and gas reservoir related problem. We also discuss about the CI and HCI models that resolved other Petroleum Engineering Problems (PEPs) since to the best of our knowledge no other significant studies have been noticed that attempted to predict the Non-HC gases in multi-stage separators using either CI or HCI models.

2.1 Application of CI models in solving PEPs

A number of studies attempted to use CI models in solving many PEPs. ANN is vastly used to resolve various PEP, e.g., PVT properties, viscosity, permeability and gas–oil ratio (GOR) prediction etc. Sundgren et al. [45] analyzed the signals from array of gas sensors with both conventional models and ANN models. They showed that hydrogen and acetone from the two-component mixture were best predicted by the ANN model. Briones et al. [8] applied ANN to predict reservoir hydrocarbon mixture components. Production performance of gas reservoirs using Type Curve is accomplished by Ameri et al. [6]. Elsharkawy and Foda [14] proposed to use GRNN to predict initial reservoir fluid compositions. They showed that the GRNN models accurately predict the changes in retrograde gas condensate composition and to estimate the pressure depletion behavior of gas condensate reservoirs. Elsharkawy [13] also used RBF network for modeling the properties of crude oil and gas system. Varotsis et al. [47] attempted a novel approach for predicting the complete PVT behavior of reservoir oil and gas condensates by using an ANN, and it is showed that the performance of ANN is more accurate than EOS models. Osman et al. [42] presented ANN model for predicting the bubble point pressure (P_b) and the oil formation volume factor (B_{ob}). ANN correlation was also developed to predict both P_b and the B_{ob} with the aid of two separate networks by Goda et al. [19]. Wong et al. [49] analyzed soft computing methods such as ANN, FL, evolutionary computing, and probabilistic reasoning approaches to address the issues of data integration, risk assessment and quantification of uncertainty in

petroleum exploration and development. Nikravesh and Aminzadeh [38] highlighted the role of soft computing, e.g., ANN, FL, GA, probabilistic reasoning in intelligent reservoir characterization. Osman and Al-Marhoun [41] proposed ANN models for predicting PVT properties of oil filed brines. Al-Farhan and Ayala [3] developed a reliable predictive tool using ANN for forecasting optimum operating conditions of a surface facility for the recovery of condensates from natural gases. Jupudi et al. [30] used the chemical percolation de-volatilization (CPD) model for predicting light gas composition. Moghadassi et al. [34] described in details about the importance of ANN against EOS and EC for estimation of PVT properties of oil/gas reservoirs. They concluded that ANN's capability to estimate the PVT properties is one of the best estimating methods with high performance. Prediction of Crude Oil Viscosity and Gas/Oil Ratio Curves Using Neural Networks are performed by Oloso et al. [40]. Elshafei et al. [12] predicted the GOR in a separator using ANN. ANN is used by Lashin and El-Din [32] to estimate and predict the most important petro-physical parameters of Nullipore reservoir based on well logging data and available core plug analyses. Eventually, as described above, ANN has been applied to predict PVT properties, viscosity, permeability, P_b, B_{ob}, etc. with varying success. It should be noted that the performance of ANN could be further enhanced by choosing appropriate ANN parameters, e.g., number of layers, nodes in the hidden layer, training algorithm.

Fuzzy Logic (FL)-based models are also used in solving various PEPs. Ali and Chawathe [4] introduced FL algorithm to rank petrographic elements with respect to their effects on permeability. Finol et al. [17] proposed a FL-based approach for permeability prediction in which the relationship between porosity and permeability was developed based on fuzzy rules. Ilkhchi [27] used fuzzy c-means clustering algorithm to classify the rock type depending on the porosity and permeability data. FL aids us with simplified knowledge acquisition and representation. Fuzzy Inference System (FIS) is a knowledge representation model where each fuzzy rule describes a local

behavior of the system and can be easily modified. However, the development of a FIS is hard to implement and require fine tuning and simulation before operational. The drawbacks of a simple FIS can be resolved when training with back-propagation algorithm results in Adaptive Neuro Fuzzy Inference System (ANFIS). ANFIS is also found very popular solving various PEPs. Permeability prediction in high water-cut reservoir is estimated using ANFIS by Sun et al. [44]. Chang and Chang [9] used ANFIS for prediction of water level in reservoir. Hurtado et al. [26] used ANFIS to obtain permeability from porosity core data. ANFIS is also used for prediction of fracture porosity of carbonate reservoir by Xiao-mei et al. [50].

SVR is also gained popularity in solving various PEPs. Nagi et al. [37] proposed SVMs for prediction in uncertain situations using the ε-Support Vector Regression (ε-SVR) method for accurate determination of the PVT properties such as the P_b and B_{ob}. Al-Anazi and Gates [1, 2] applied SVM technique for classifying electrofacies of a highly heterogeneous sandstone reservoir. They also successfully predict permeability distributions by using SVR. Khoukhi et al. [31] proposed SVMs for viscosity and gas/oil ratio curves prediction. It is evident that the performance of SVR depends on the kernel function, regularization parameter C and epsilon (ε), the control parameters of support vectors. These important parameters can be optimized using GA and thus overcome the performance and complexity of a simple SVR.

2.2 Application of HCI models in solving PEPs

The trend of using HCI models is also appeared in solving various PEPs and proved successful in literatures. The evolutionary algorithm is adopted to ensure auto-design of CI models without any intervention from the designer. A hybrid Neuro Genetic approach is attempted for hydraulic fracture treatment design and optimization by Mohaghegh et al. [35]. In a later study Mohaghegh et al. [36] proposed a method for selection of maximum number of candidate well for stimulation of gas storage wells using Neural Networks and Genetic Algorithms (GAs). Saemi et al. [43] addressed the limitations of traditional design and structural optimization process of ANN. They proposed an evolutionary ANN approach to predict permeability using well log data. They have used GAs to determine the number of neurons in the hidden layers, the momentum and the learning rates for minimizing the time and effort required to find the optimal architecture. In this paper, we define the number of neurons in the hidden layer as well as the hidden layer's activation function and the training algorithm that suits the structure. Oloso et al. [39] proposed a genetically optimized ANN for predicting viscosity and gas/oil ratio curves. Prediction of hydrocarbon reservoir

characteristics using GA-based optimized RBF neural network are introduced by Chen et al. [10]. In this study, unlike others, we tried to optimize the structure of the network while optimizing the better training algorithm as well as suitable activation function in the hidden layer.

GA-based Neuro-Fuzzy models are found in few related studies. Xie et al. [51] developed a hybrid genetic programming and Fuzzy/NN inference system to estimate permeability for all rock types or lithologies by utilizing lithologic and permeability facies as indicators. They showed that the hybrid model is robust in estimating permeability in complex heterogeneous reservoirs. Ghouti and Bukhitan [18] proposed prediction of both P_b and B_{ob} using Hybrid of Neuro-Genetic Fuzzy Inference System.

GA-based SVR are applied in solving various PEPs. Fei et al. [16] proposed SVM with GA to forecast the ratios of key-gas in power transformer oil. Helmy et al. [24] proposed SVM and FL driven HCI models for characterization of oil and gas reservoirs. In a different study, Helmy and Fatai [23] also used SVM and FL driven HCI models for porosity and permeability prediction of petroleum reservoirs. The problem that we are dealing with is more challenging than the problems solved by the above studies.

To the best of our knowledge, no noticeable work has been carried out in the field of gas composition prediction in multistage separator using HCI models. Hossain et al. [25] investigated the prediction of Non-HCs in separator using ANN, and the results show prominent accuracy of prediction. They showed that prediction accuracy of ANN is comparable with the conventional methods which motivate us to investigate the application of optimized HCI models for predicting the Non-HC gas components. Hence, the novelty of our work lies in twofold: application of HCI to solve the problem of predicting the Non-HC gases of multi-stage separator and enhance the accuracy for the prediction using GA. In the industry, EOS and EC are used to estimate oil and gas properties. We have used Peng–Robinson-based EOS (PR-EOS) which takes reservoir gas compositions, C7+ molecular weight (MW) and density at reservoir, separator stage temperature and pressure as inputs to calculate the gas compositions on that stage. In this paper, we compared the performance of the CI and HCI models' with the performance of PR-EOS.

3 Fundamentals of CI models

3.1 Artificial Neural Network

ANN is a CI model inspired by the way in which the brain performs a particular learning task. Multi-layer perceptron (MLP) is one of the most popular ANNs that has gained vast popularity in many research areas including petroleum

engineering [3, 6, 34, 45]. MLP has one input layer, one output layer, and one or more hidden layers of processing units. It can be trained to perform a particular function by adjusting the values of the connections (weights) between a particular input and the specific target output.

3.2 Support Vector Regression

SVM proposed by Boser et al. [7] which is based on the basis of statistical learning theory. SVR is a regression version of SVMs. The main idea is the same for both SVR and SVM which is to minimize error and individualizing the hyper-plane which follows the maximum margin algorithm: a nonlinear function is learned by linear learning machine mapping into high-dimensional kernel-induced feature space (φ) (Fig. 2). Margin is a distance between optimal hyper-plane and the support vectors that lies closer to it (Fig. 2). Figure 2a shows the two dimensional data having the circular decision boundary which is linearly non-separable. Figure 2b shows the mapping of the data into three dimensional spaces where the circular decision boundary becomes a linear hyper-plane. Figure 2c shows the two dimensional projection of the Fig. 2b. To explain this mathematically, considering $x \in R^n$ and $y \in R$. The hyper-plane function is presented as $y = f(a) = w \cdot \varphi(x) + b$ where $w \in R^n$ is the weight vector, $b \in R$ is the bias. The function $\varphi(x)$ denotes a nonlinear transformation from R^n into a higher-dimensional space. It is aimed to find the value of w and b such that values of x can be determined by minimizing the regression risk

$$R = \frac{1}{2} \sum_{i=1}^{n} \{f(x_i) - y_i\}^2 + \frac{\lambda}{2} \|w\|^2$$

where n indicates the sample inputs $(x_1, \ldots, x_n)$, λ is regularization constant, $(y_1, \ldots, y_n)$ are the measurements.

SVM/SVR is successfully applied in various research areas including petroleum, e.g., [1, 2, 31, 37]. The computational complexity of an SVM/SVR depends on the number of support vectors used in the procedure [46]. In this research we have used the SVR model that trained with reduced number of support vectors by using K-means algorithm known as SVM-KM [5, 46, 52].

3.3 Adaptive Neuro Fuzzy Inference System

ANFIS is introduced by Jang [28] which serves as a basis for constructing a trained FIS. The Neuro-Fuzzy methods provide models in the form of the "IF–THEN" rules that can be easily interpreted by human beings. The structure of ANFIS consists of five-layered feed forward network. A typical example of an ANFIS model is shown in Fig. 3. In this hybrid model, the consequent parameters of Takagi–

Sugeno–Kang (TSK) fuzzy inference model [28] is trained through least-square method and the antecedent parameters are obtained with the help of gradient-descent method. Figure 3a shows the reasoning mechanism that can be implemented into a feed-forward neural network with supervised learning capability, which is known as ANFIS architecture (Fig. 3). Jang et al. developed a hybrid-learning rule for ANFIS which is faster than the classical back-propagation method by combining the gradient method and the least squares estimate to identify antecedent and consequent parameters. The square nodes in Fig. 3 indicate adaptive nodes with parameters and circle modes indicate fixed nodes without parameters. ANFIS basically implements a first-order Sugeno-style fuzzy system. Although it is quite easy to express linguistically the relation between input and output, it is difficult to fit the fuzzy model to the target data using trial and error. A better approach is to approximate the target function with a piece-wise linear function and interpolate it, in some way, between the linear regions. In the TSK model the idea is that each rule in a rule base defines a region for a model, which can be linear. This is achieved by clustering the input space. We have used subtractive clustering (Subclust) to create initial FIS and then trained that FIS using ANFIS hybrid learning algorithm. The functionality of nodes in ANFIS (Fig. 3), as a five layered feed-forward neural structure layers can be summarized as follows. *Layer 1* consists of square nodes that perform fuzzification with chosen membership. In *Layer 2* the T-norm operation is performed to produce the firing strength of each rule. In *Layer 3*, the nodes are fixed with outputs generating the normalized firing strengths by calculating the ratio of the ith rule firing strength to the sum of all rules' firing strength is calculated in the third layer. *Layer 4* consists of square nodes that perform multiplication of normalized firing strengths with the corresponding rule. The parameters in this layer are called consequent parameters. In *Layer 5*, the single node is fixed with output which calculated by the sum of all incoming signals. A Two Rule Sugeno ANFIS of Fig. 3 has rules of the form:

If x is p_1 and y is q_1 THEN $f_1 = p_1 x + q_1 y + r_1$

If x is p_2 and y is q_2 THEN $f_1 = p_1 x + q_2 y + r_2$

The ANFIS architecture was successfully used to model various PEPs [9, 26, 44, 50].

3.4 Genetic Algorithm

GA is a global stochastic search techniques based on biological evolution introduced by John Holland in the 1970s. Robustness of GA in solving nonlinear optimization problems are discussed later by Goldberg [20]. GA is used to determine the global optima or the sub-optima of a given
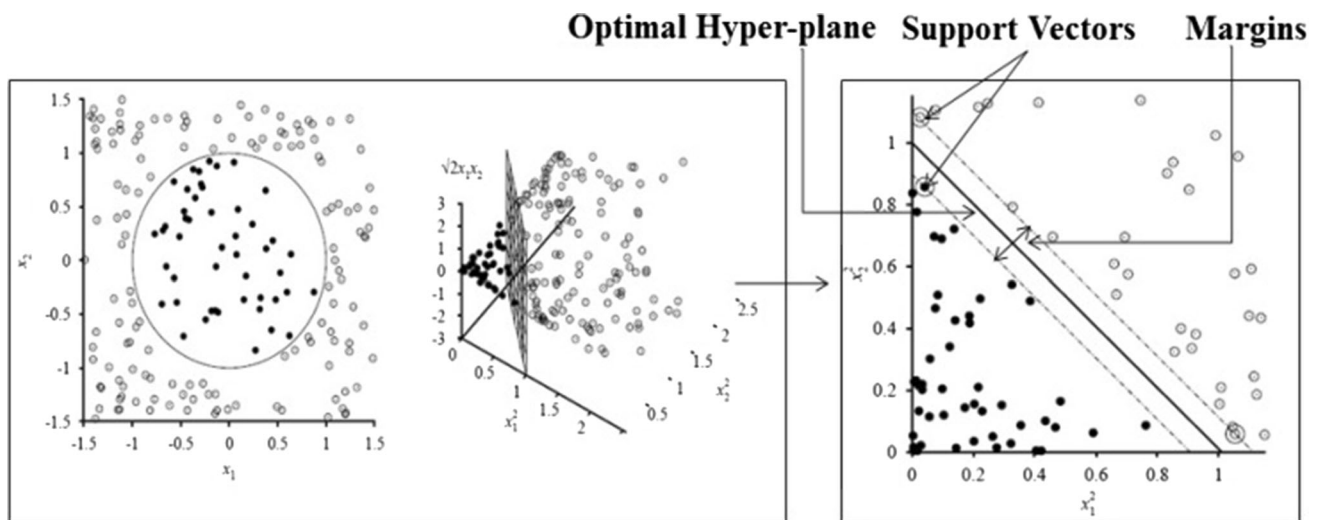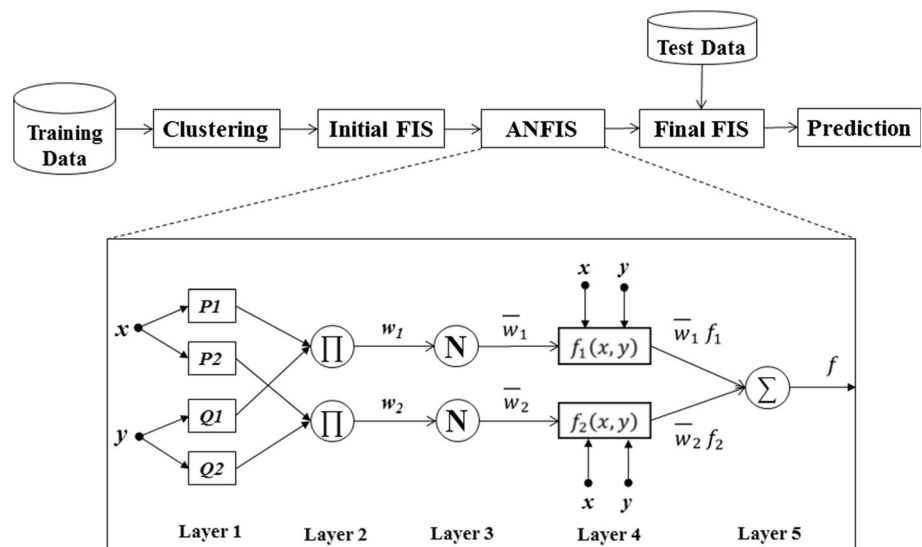
Fig. 2 Mapping of input vectors in higher dimension in SVM

Fig. 3 Train and test steps of ANFIS and its structure

function or a process that may subject to constraints. GA starts with a randomly generated population where the variables are represented as a string of elements known as chromosome. A set of "genetic" operators such as Selection, Crossover and Mutation is then applied to this population set so as to create a new population set in search of the optimal solution. The selection operator chooses chromosomes from the current population based on fitness value of the individuals. The crossover operator combines the features of two parent chromosomes to form two similar offspring by swapping corresponding segments of the parents. The mutation operator creates new chromosomes by randomly changing the genes of existing chromosomes. GA can explore the entire design space by the genetic manipulations. It does not easily fall into a certain local minima or maxima by maintaining a balance between the

two conflicting objectives of a search procedure, i.e., the exploitation of the best solution and the exploration of the search space. Therefore, GA is an aggressive search technique that quickly converges to find the optimal solution in a large solution domain. The GA was successfully used to model various PEPs [11, 22, 33].

4 Proposed HCI models for Non-HC gases prediction: GA−MLP, GA−SVR and GA−ANFIS

The accuracy of the CI models depends on their parameters for a given set of data. Finding the optimal parameters of the CI models is a big challenge. The parameters can be chosen by trial and error basis which is not actually

Table 1 GA-optimized parameters of CI models for Non-HCs prediction

Non-HC	GA-optimized parameters					
	MLP		ANFIS	SVR		
	# Nodes	Act Fn	Radius	C	Lambda (λ)	Epsilon (ε)
N_2	56	log	0.2998	0.9763	0.00067	0.1754
CO_2	21	tan	0.612	6.2202	0.0005	0.0352
H_2S	17	log	0.7959	0.8796	0.00044	0.0001

efficient and time consuming. Furthermore, trial and error basis do not guaranty an optimal or sub-optimal solution. A stochastic search technique is a better alternative to search the optimal/sub-optimal parameters of the CI models in a given search space. GA-optimized CI models are applied successfully in various petroleum studies includes forecast the ratios of key-gas in power transformer oil [16], characterization of oil and gas reservoirs [24], porosity and permeability prediction [23].

In this paper, the crucial parameters of ANN, SVR and ANFIS are optimized using GA. To find the optimal/sub-optimal parameters of CI models, 20 generations having 50 populations in each generation and 0.65 crossover fraction are used as GA parameters. The objective function is designed to achieve the optimal parameters for which the CI models are having the minimum root-mean-square error (RMSE). The CI models are trained and experimented on the Non-HC gases in the separator of an GOSP. The Fig. 4 depicts the generic flow chart of the parameter optimization of CI models using GA. As shown in Fig. 4, the accuracy of the CI models depends on their parameters for a given set of data. Finding the optimal parameters of the CI models is a big challenge. The parameters can be chosen by trial and error basis which is not actually efficient and time consuming. Furthermore, trial and error basis do not guaranty an optimal or sub-optimal solution. A stochastic search technique is a better alternative to search the optimal/sub-optimal parameters of the CI models in a given search space. Performing optimization by GA with different fixed parameters results into a completely different architecture of the CI model in each run. The performance of the HCI model mostly depends on the design of the objective function. We have used RMSE as criterion of measuring fitness (Eq. 1). We have designed the objective functions where in the validation dataset of samples n is having minimum RMSE.

$$\text{Fitness(RMSE)} = \sqrt{\frac{(x_n - y_n)^2}{n}} \qquad (1)$$

4.1 Ga−MLP

In this paper we have optimized three crucial MLP parameters such as number of hidden neurons in the hidden layer, activation function of the hidden layer and the training algorithm. As we have a limited number of training samples, we decided to keep the structure of MLP small and use only one hidden layer. Additionally, we kept the range of hidden nodes to be optimized between 1 and 63 to keep the network simpler. While performing the initial runs, we observed that the results were fluctuating among the log and tan sigmoidal activation functions. At the same time, Levenberg–Marquardt (LM) training algorithm was providing comparable results with resilient back-propagation (Rprop). Therefore, we also decided to optimize the suitable activation function in the hidden layer and the type of the training algorithm that would fit that structure using GA. We found that the training algorithm LM performs better than the Rprop for all three Non-HCs. On the other hand, log sigmoidal activation function in the hidden layer is well fitted with N_2, H_2S and tan sigmoidal activation function fitted with CO_2. The other MLP parameters that we kept fixed are the learning rate 0.001, epochs 300 and error goal 0.00001. As we have predicted one gas component at a time, we have used one node in the output layer. We varied the output between −1 and +1 and so used tan sigmoidal activation function in the output layer. Figure 5 shows the input and the output parameters that fed into the MLPs to train the model to perform prediction of N_2, H_2S and CO_2 separately. Figure 6a depicts the generic flowchart of the parameter optimization of MLP using GA. Table 1 shows the optimal parameters of MLP obtained through GA.

4.1.1 Chromosome encoding of GA−MLP

We have used binary encoding to represent chromosome for MLP parameters optimization. The chromosome of MLP includes bit string of 8 digits Fig. 6a where the right most two bits were used to choose the training algorithm and the activation function of the hidden nodes. And the rest 6 digits were used to for finding the optimal network structure, i.e., the number of hidden nodes in the hidden layer.

4.2 Ga−SVR

The optimal parameter search on SVR plays a crucial role in building a prediction model with high prediction accuracy and stability. From the initial run of the problems, we have decided to use "polynomial" type kernel having degree of polynomial 0.5. GA can
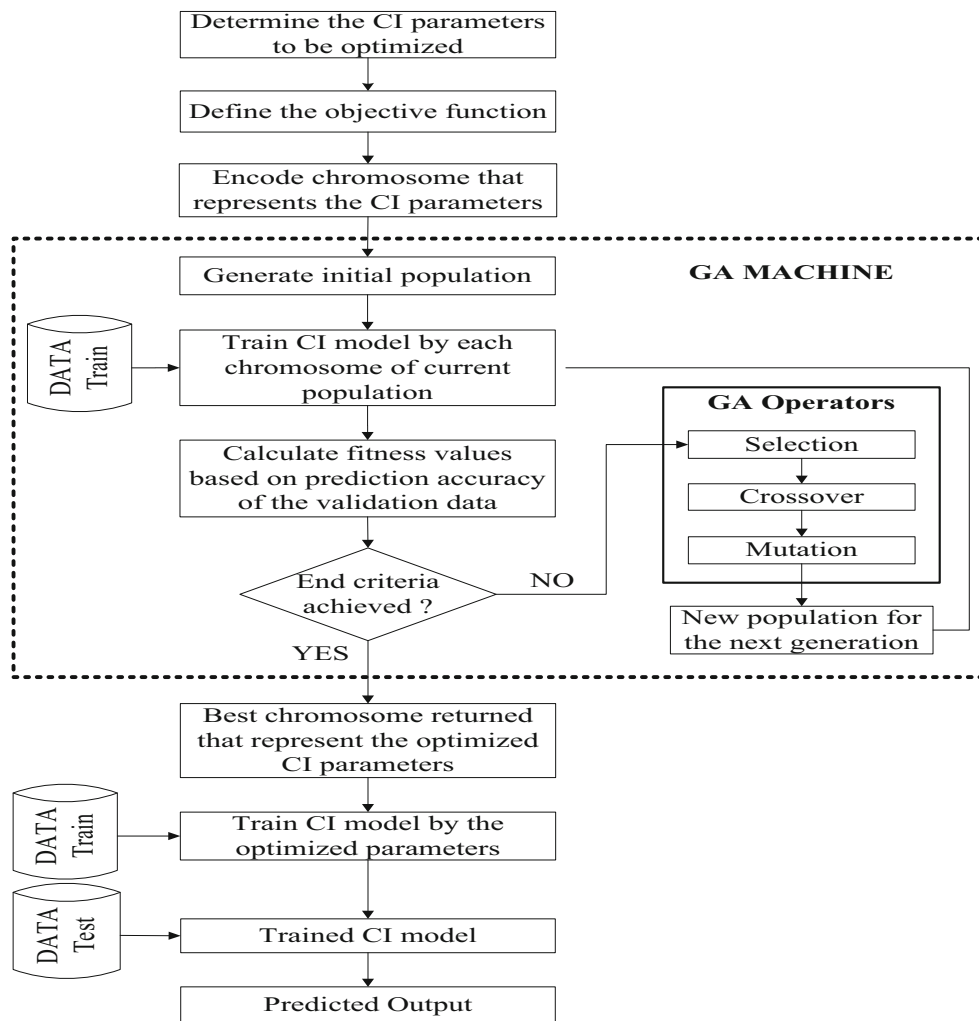
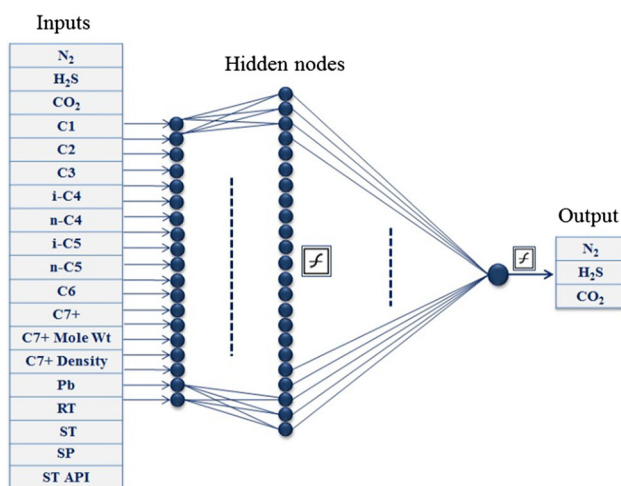**Fig. 4** Flow chart for optimization of CI models using GA

Fig. 5 Structure of MLP with input and output parameters for Non-HC gas prediction

automatically optimize the SVR parameters C, epsilon (ε) and lambda (λ) and thus increase the predictive accuracy and capability of generalization. The chromosome for SVR is encoded into real-valued encoding in the following ranges of C (0.0001–100), ε (0.0001–0.6), λ (0.000000001–0.001). The Fig. 6b depicts the generic flow chart of the parameter optimization of SVR using GA. The optimum parameters of SVR that are obtained through GA are shown in Table 1.

4.2.1 Chromosome encoding of GA−SVR

We have used real-valued encoding to represent chromosome for SVR parameters optimization. The chromosome of SVR includes the three real numbers representing C, epsilon, lambda (Fig. 6b).
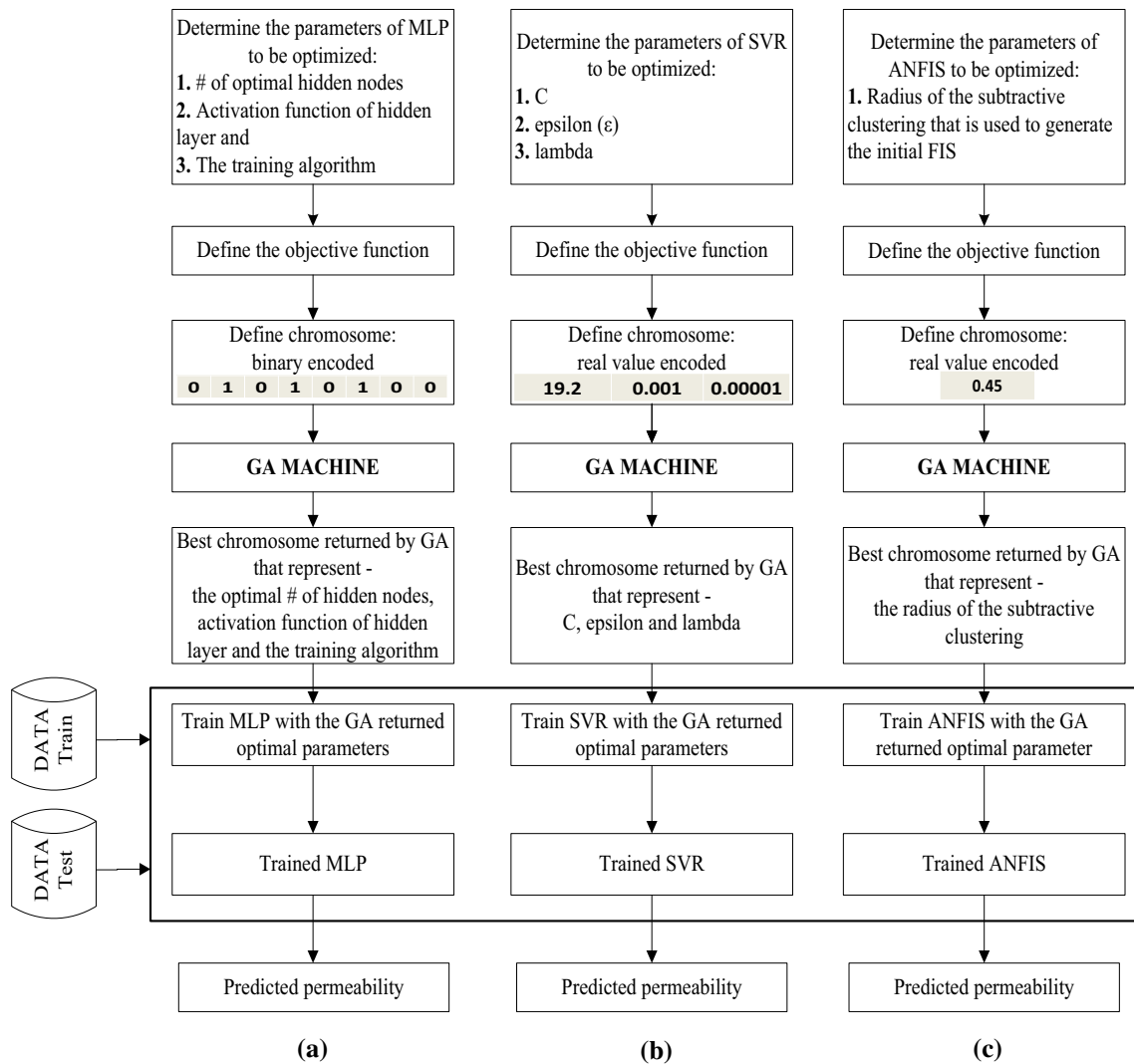
Fig. 6 Flowcharts for GA-optimized CI models: **a** GA−MLP, **b** GA−SVR, **c** GA−ANFIS

4.3 Ga−ANFIS

The performance of ANFIS depends on the initial FIS. The more the initial FIS represented better, the better the performance of ANFIS would be. We have created the initial FIS by using Subclust. Subclust is one of the clustering algorithms based on a measure of the density of data points in the feature space. It generates the rules that approximate a function. The rule extraction method first uses Subclust to determine number of rules and input membership functions equation. We decided to optimize the radius of Subclust by GA. The range of the radius we choose to be between 0.2 and 0.9. Figure 6c depicts the flow chart of the parameter optimization of ANFIS using GA. The optimum parameter that obtained through GA is shown in Table 1.

4.3.1 Chromosome encoding of GA−ANFIS

We have used real-valued encoding to represent chromosome for ANFIS parameters optimization. In this case, it is only one real value (Fig. 6c) that represents the radius of the Subclust used to create the initial FIS to be trained in ANFIS training phase.

5 Data and experimental setup

In this paper, around 80 reservoir crude oil compositions and the corresponding separator gas compositions are collected from various Asian reservoirs and related studies [29]. Our goal is to predict the Non-HC gas components in the separator at certain stage temperature and pressure from the

Table 2 Statistical characteristics of the data samples used for Non-HCs gas prediction models

Parameters	Min	Max	Mean	SD
Input parameters				
N_2	0.00	4.70	0.35	0.76
CO_2	0.00	7.38	1.69	2.19
H_2S	0.00	12.37	0.77	1.84
C1	0.52	47.70	19.00	13.98
C2	1.29	14.05	7.13	3.08
C3	2.83	12.03	6.86	1.66
i_C4	0.79	3.34	1.42	0.60
n_C4	2.45	7.76	4.48	1.05
i_C5	1.13	4.22	1.97	0.69
n_C5	1.38	4.79	2.92	0.77
C6	1.88	9.79	4.36	1.64
C7+	23.06	75.11	49.08	15.18
C7+ density	0.72	0.93	0.87	0.03
C7+ MW	193.00	350.00	261.15	38.84
ST	50.00	315.00	117.38	33.65
SP	14.70	519.00	135.20	113.56
P_b	381.00	3986.00	2172.07	754.31
ST API	24.20	124.10	36.99	10.05
RT	130.00	280.00	193.71	27.32
Output parameters				
N_2	0.00	0.73	1.56	9.66
CO_2	0.00	3.80	4.35	16.44
H_2S	0.00	1.17	2.82	18.86

available reservoir crude oil compositions and other reservoir parameters. The training input parameters of the CI models consist of mole percent of the Non-HC, e.g., N_2, H_2S and CO_2 and HC, e.g., methane (CH_4 as C1), ethane (C_2H_6 as C2), propane (C_3H_8 as C3), butane (C_4H_{10} as C4), pentane (C_5H_{12} as C5), hexane (C_6H_{14} as C6), heptane and heavier (C7+) contents of reservoir crude oil sample. The isomers of C4 and C5 are counted separately. The other available reservoir parameters that we have used as input to train the CI and HCI models are Stock Tank American Petroleum Institute (ST API) gravity, P_b, reservoir temperature (RT), separator pressure (SP) and separator temperature (ST). The pressures are measured in unit "psi" and temperatures in unit "°F". MW and density parameters of C7+ compositions are also given as input parameters. In this paper, we focused on the prediction of Non-HC consist of mole percent of N_2, CO_2 and H_2S. As we have only 80 reservoirs samples, we have selected random 10 samples for testing and the remaining samples used for training. Statistical characteristics of the data samples used for Non-HCs gas prediction are shown in Table 2. For the ANN, we have used learning rate: 0.001, epochs: 300, error goal: 0.00001. The other parameter we decided to be optimized by GA. We tried to

achieve the number of nodes in the hidden layer. We kept the range of hidden nodes between 1 and 63 to keep the network simpler. The second parameters that we achieved by GA is the hidden neurons' activation function. The possible options for activation functions are either tan-sigmoid or log-sigmoid. We have also found the training algorithm either LM or Rprop (Resilient Back propagation). We have used binary encoding to represent chromosome for MLP parameters optimization. We have used population type bit string. And the right most bit is to decide the training algorithm. The second bit is to choose activation function in the hidden layers. And the left most 6 bits is to choose the number of nodes in the hidden layer. We have selected different sizes of training set for different types of output. At first we divided the whole datasets into training and testing. Around 80 % of the whole datasets is used for training and the 20 % of the relevant datasets were used for testing.

6 Performance evaluation

The commonly used techniques to evaluate the performance of prediction in the related work are the RMSE and the Correlation Coefficient (CC). Additionally, a good prediction model should have significant level (p value) within 5 %. A p value represents the probability of finding a co-relation by chance. In terms of statistical significance, the lower the p value, the less likely the result is if the null hypothesis is true, and consequently the more "significant" the result is. The null hypothesis is often rejected when the p value is less than 0.05 or 0.01. Furthermore, we have calculated the training time of the models which is computed by deducting the end CPU time of training from the beginning of the training CPU time.

6.1 Performance metric

We have used box metric to represent CC and RMSE so as to easily compare the outcomes of the models. We gave most importance to error measure of a model that is the RMSE values as long as it has an accepted CC value. In statistics CC value greater than 0.75 represents strong correlation between the predicted output and original values.

7 Experimental results and discussions

In this paper we have shown the performance of each model in the Fig. 8a–f. Figure 8a, c, e depicts the performance in terms of the metric CC versus RMSE and Fig. 8b, d, f shows the regression analysis of the prediction of testing data for the best performed model. Figure 7 shows the training time taken by the CI and HCI models for Non-HCs gas predictions in multi-stage separator. Tables 3, 4, 5
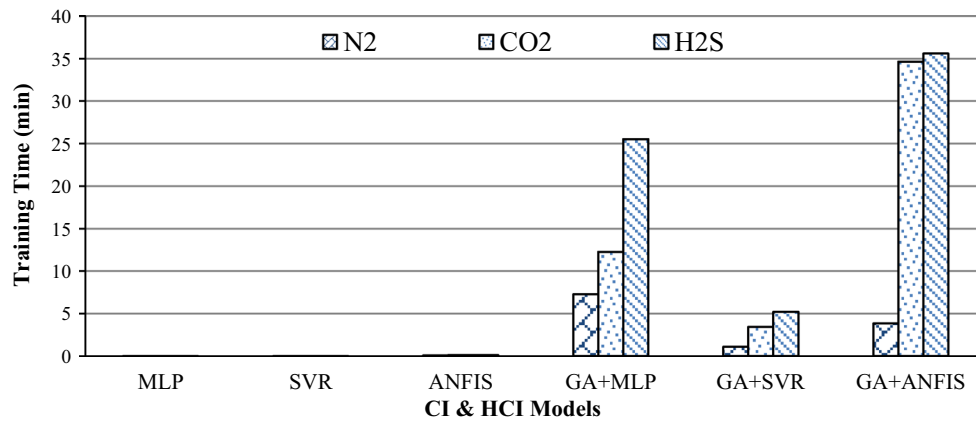
Fig. 7 Training time of CI and HCI models for Non-HCs gas prediction

Table 3 Performance of CI models on training data

Training	PR-EOS		MLP			SVR			ANFIS		
Non-HC	CC	RMSE	CC	p value	RMSE	CC	p value	RMSE	CC	p value	RMSE
N_2	0.8008	1.1383	0.9233	0.0000	0.8834	0.8142	0.0000	1.2879	0.9426	0.0000	0.5110
CO_2	0.9978	0.2926	0.9910	0.0000	0.6219	0.9865	0.0000	2.3149	0.9976	0.0000	0.2927
H_2S	0.9947	0.5334	0.9908	0.0000	0.6097	0.9799	0.0000	2.9719	0.9944	0.0000	0.3241

Table 4 Performance of CI models on test data

Test	PR-EOS		MLP			SVR			ANFIS		
Non-HC	CC	RMSE	CC	p value	RMSE	CC	p value	RMSE	CC	p value	RMSE
N_2	0.9586	0.7402	0.9403	0.0001	1.5183	0.9843	0.0000	1.2947	0.9764	0.0000	0.6331
CO_2	0.9989	0.3114	0.9669	0.0000	1.2819	0.9419	0.0000	2.0379	0.9693	0.0000	1.1854
H_2S	0.6800	2.0037	0.9160	0.0019	1.2807	0.9149	0.0002	2.8769	0.7828	0.0044	1.0078

Table 5 Performance of HCI models on training data

Training	PR-EOS		GA + MLP			GA + SVR			GA + ANFIS		
Non-HC	CC	RMSE	CC	p value	RMSE	CC	p value	RMSE	CC	p value	RMSE
N_2	0.8008	1.1383	0.9239	0.0000	0.6173	0.7954	0.0000	0.9110	0.8965	0.0000	0.6655
CO_2	0.9978	0.2926	0.9932	0.0000	0.5030	0.9890	0.0000	0.6285	0.9991	0.0000	0.1845
H_2S	0.9947	0.5334	0.9786	0.0000	0.7012	0.9963	0.0000	0.2969	0.9990	0.0000	0.1388

and 6 show the numerical values of the CC, RMSE and p values of all the models. The CC value represents how good the prediction is and the p value shows how significant the prediction is. The CC above 0.75 represents statistically acceptable correlation, and the p value less than or equal to 0.05 means the significance level is within 5 %. In Tables 3, 4, 5 and 6, we can see that the p value is less than 0.05 except one or two cases which shows that the prediction of models are significant.

7.1 Nitrogen (N_2)

In case of N_2 prediction, the performance of two HCI and one CI models is better than the PR-EOS-based model. N_2 is best predicted by genetically optimized ANFIS model (Fig. 8a). Figure 8b shows the regression analysis of the GA + ANFIS model on the test data. The performance of only ANFIS is also close to the optimized ANFIS. The performance of GA + SVR is closer to the PR-EOS-based

Table 6 Performance of HCI models on test data

Test	PR-EOS		GA + MLP			GA + SVR			GA + ANFIS		
Non-HC	CC	RMSE	CC	p value	RMSE	CC	p value	RMSE	CC	p value	RMSE
N_2	0.9586	0.7402	0.8953	0.0005	1.1926	0.9767	0.0000	0.7183	0.9795	0.0000	0.5851
CO_2	0.9989	0.3114	0.9920	0.0000	0.6492	0.9385	0.0001	1.6950	0.9827	0.0000	0.9162
H_2S	0.6800	2.0037	0.8527	0.0017	0.9093	0.9119	0.0002	0.7088	0.8211	0.0036	0.9878

method though having higher CC hence more accurate. It can be observed from Fig. 8a is that the HCI models are performed better than their corresponding simple CI models. The GA + SVR are having least training time comparing to other models (Fig. 7). On the other hand, the GA + ANFIS having the highest training time though having highest prediction accuracy. It can be observed from Fig. 7a that the HCI models are performed better than their corresponding simple CI models. The GA + SVR model is having least training time comparing to other HCI models. On the other hand, the GA + ANFIS model has the highest training time among all the models though having highest prediction accuracy in case of N_2.

7.2 Carbon dioxide (CO_2)

In case of CO_2 the performance of PR-EOS-based model is better than both the CI and HCI models (Fig. 8c). This is may be because we do not perform post processing so that very small value counted as negative. We believe that post processing of the output can improve the performance of CO2 prediction by eliminating negative predicted value. Among the CI and HCI models, the performance of GA + MLP is better than any other models. The regression analysis of the GA + MLP model in Fig. 8d also shows good co-relation among the test data with the prediction. On the other hand, the performance of GA + ANFIS is comparable with GA + MLP. Additionally, Fig. 8c clearly shows the performance of HCI models is better than the corresponding CI models. Figure 7 stated that the GA + MLP model took highest training time. The statistics of Table 2 shows that the Non-HC component output parameter CO_2 varies in wide ranges. This could be a reason that empirical PR-EOS model performs better than the CI and HCI models. Another reason could be the less error-prone behavior of PR-EOS model to CO_2 calculation. On the other hand, we came up with CI and HCI models that perform in general better for the two Non-HCs components (N_2 and H_2S). The point to be noted is, the performance of PR-EOS is static and unable to improve further, whereas the development of CI and HCI models is open and further improvement of CI and HCI models are

still possible by using different CI models and optimization techniques.

7.3 Hydrogen sulfide (H_2S)

In case of H_2S the performance of CI and HIC models is better than the PR-EOS model except SVR (Fig. 8e). In contrast the GA + SVR outperformed all others models. On the one hand, it is performing with lowest RMSE while having highest co-relation (Fig. 8e) and on the other hand, it took the least training time (Fig. 7). Among the other models the performance of GA + ANFIS and GA + MLP is comparable and closer to GA + SVR. The training time of GA + MLP and GA + ANFIS is huge comparing other models (Fig. 7). Figure 7 shows that the HCI models took huge training time comparing simple CI models though having higher accuracy.

8 Conclusion and future work

In this paper, Non-HC gas components prediction in multi-stage separator is accomplished by using various CI and HCI models. Choosing optimal parameters of the CI models is a big challenge. GA is used to find the optimal parameters in the proposed HCI models. The results clearly show that the proposed HCI models predict Non-HC gas components in multi-stage separator with higher accuracy than the conventional PR-EOS-based method for most cases. We observed that the performance of both CI models and HCI models are better than the performance of conventional PR-EOS in predicting N_2 and H_2S. Though the PR-EOS-based method performs well in predicting CO_2, the results obtained by the CI and HCI models are comparable. It is observed that different CI and HCI models are suitable for different Non-HC. For example, GA + ANFIS performed well for N_2 prediction, GA + MLP for CO_2 and GA + SVR for H_2S. Therefore, one particular HCI model cannot be recommended to predict all the gas components. As a future work and to improve the accuracy of prediction, different combination of CI or HCI models with additional data including feature selection methods can be
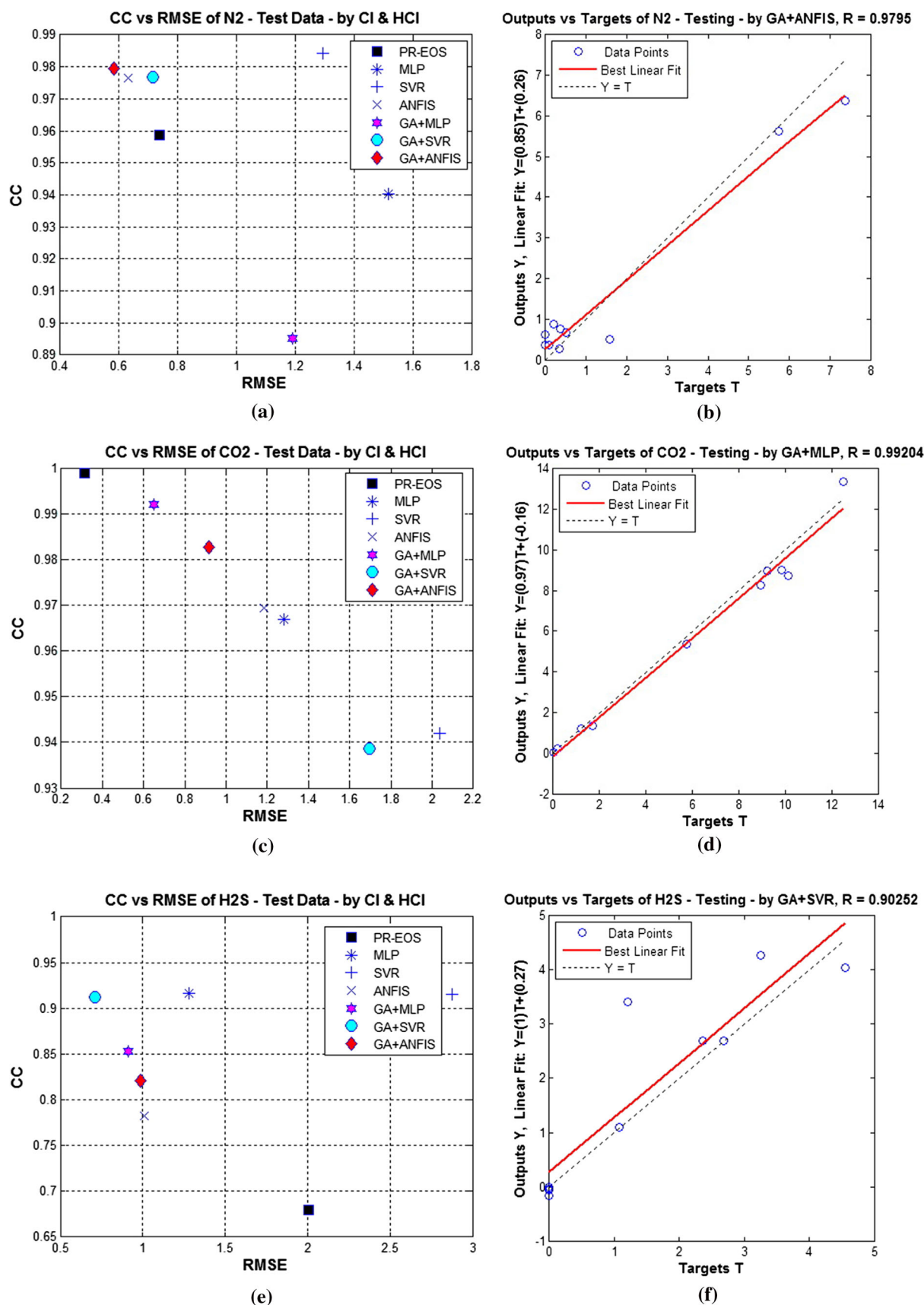
Fig. 8 Performance of CI and HCI models for Non-HC gas predictions in multi-stage separator

studied to select the appropriate models for predicting every gas components in multi-stage separator.

References

1. Al-Anazi A, Gates ID (2010) A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. Eng Geol 114(3–4):267–277

2. Al-Anazi A, Gates ID, Azaiez J (2009) Innovative data-driven permeability prediction in a heterogeneous reservoir. In: SPE EUROPEC/EAGE annual conference and exhibition, Amsterdam, The Netherlands

3. Al-Farhan FA, Ayala LF (2006) Optimization of surface condensate production from natural gases using artificial intelligence. J Petrol Sci Eng 53(1–2):135–147

4. Ali M, Chawathe A (2000) Using artificial intelligence to predict permeability from petrographic data. Comput Geosci 26:915–925

5. Almeida MB, Braga AP, Braga JP (2000) SVM-KM: speeding SVMs learning with a priori cluster selection and k-means. In: Proceedings of the 6th Brazilian symposium on neural networks, pp 162–167

6. Ameri S, Aminian K, Mohaghegh S (1994) Predicting the production performance of gas reservoirs. Sci Iran 1(3):231–240

7. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Haussler D (ed) Proceedings of the 5th annual ACM workshop on computational learning theory, pp 144–152

8. Briones MF, Corpoven SA, Rojas GA, Martinez ER (1994) Application of neural network in the prediction of reservoir hydrocarbon mixture composition from production data. In: SPE annual technical conference and exhibition, New Orleans, Louisiana

9. Chang F, Chang Y (2006) Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Adv Water Resour 29(1):1–10

10. Chen J, Li Z, Zhao D (2009) Prediction of hydrocarbon reservoir parameter using a GA-RBF neural network. In: Computational intelligence and intelligent systems, 4th international symposium, ISICA 2009, Huangshi, China, pp 379–386

11. Monyei CG, Adewumi AO, Obolo MO (2014) Oil well characterization and artificial gas lift optimization using neural networks combined with genetic algorithm. Discrete Dyn Nat Soc 2014:289239. doi:10.1155/2014/289239

12. Elshafei M, Khoukhi A, Abdulraheem A (2010) Neural network aided design of oil production units. In: 2010 10th International conference on information sciences signal processing and their applications (ISSPA). IEEE, pp 638–641

13. Elsharkawy AM (1998) Modeling the properties of crude oil and gas systems using RBF network. In: SPE Asia Pacific oil and gas conference, Perth, Australia

14. Elsharkawy AM, Foda SG (1998) EOS simulation and GRNN modeling of the constant volume depletion behavior of gas condensate reservoirs. Energy Fuels 12(2):353–364

15. Anifowose FA, Labadin J, Abdulraheem A (2013) Prediction of petroleum reservoir properties using different versions of adaptive neuro-fuzzy inference system hybrid models. Int J Comput Inf Syst Ind Manag Appl 5:413–426

16. Fei S, Liu C, Miao Y (2009) Support vector machine with genetic algorithm for forecasting of key-gas ratios in oil-immersed transformer. Expert Syst Appl 36(3):6326–6331

17. Finol JJ, Guo YK, Jing XDD (2001) Permeability prediction in shaly formations: the fuzzy modeling approach. Geophysics 3(67):817–829

18. Ghouti L, Bukhitan S (2010). Hybrid soft computing for PVT properties prediction. In: European symposium on artificial neural networks, computational intelligence and machine learning (ESANN), Belgium

19. Goda HM, Fattah KA, Shokir EM, Sayyouh MH (2003) Prediction of the PVT data using neural network computing theory. In: 27th Annual SPE international technical conference and exhibition, Abuja, Nigeria

20. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Reading. ISBN 978-0201157673

21. Hallman JH, Cook I, Muqeem MA, Jarrett CM, Shammari HA (2007) Fluid customization and equipment optimization enable safe and successful underbalanced drilling of high-H_2S horizontal wells in Saudi Arabia. In: IADC/SPE managed pressure drilling and underbalanced operations conference and exhibition, Galveston, Texas, USA

22. Nooruddin H, Anifowose F, Abdulazeez A (2014) Using soft computing techniques to predict corrected air permeability using Thomeer parameters, air porosity and grain density. Comput Geosci 64:72–80

23. Helmy T, Fatai A (2010) Hybrid computational intelligence models for porosity and permeability prediction of petroleum reservoirs. Int J Comput Intell Appl 9(4):313–337

24. Helmy T, Fatai A, Faisal KA (2010) Hybrid computational models for the characterization of oil and gas reservoirs. Expert Syst Appl 37:5353–5363

25. Hossain MI, Helmy T, Hassan MR, Adbulraheem A, Khoukhi A, Elshafei M (2012) Non-hydrocarbons gas components prediction in multistage separator using neural networks. In: Global conference on power control and optimization PCO, Dubai

26. Hurtado N, Aldana M, Torres J (2009) Comparison between neuro-fuzzy and fractal models for permeability prediction. Comput Geosci 13:181–186

27. Ilkhchi K (2006) A fuzzy logic approach for the estimation of permeability and rock types from conventional well log data: an example from the Kangan reservoir in Iran Offshore Gas Field. Iran J Geophys Eng 3:356–369

28. Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685

29. Jaubert JN, Avaullee L, Souvay JF (2002) A crude oil data bank containing more than 5000 PVT and gas injection data. J Petrol Sci Eng 34(1–4):65–107

30. Jupudi RS, Zamansky V, Fletcher TH (2009) Prediction of light gas composition in coal devolatilization. Am Chem Soc Energy Fuels 23:3063–3067

31. Khoukhi A, Oloso M, Abdulraheem A, El-Shafei M, Al-Majed A (2011) Support vector machines and functional networks for viscosity and gas/oil ratio curves prediction. Int J Comput Intell Appl 10(3):269–293

32. Lashin A, El-Din SS (2012) Reservoir parameters determination using artificial neural networks: Ras Fanar field, Gulf of Suez, Egypt. Arab J Geosci 6(8):1866–7511

33. Mahmudi M, Sadeghi MT (2013) The optimization of continuous gas lift process using an integrated compositional model. J Petrol Sci Eng 108:321–327

34. Moghadassi AR, Parvizian F, Hosseini SM, Fazalali AR (2009) A new approach for estimation of PVT properties of pure gases based on artificial neural network model. Braz J Chem Eng 26(1):199–206

35. Mohaghegh S, Balan B, Ameri S, McVey DS (1996) A hybrid neuro-genetic approach to hydraulic fracture treatment design and optimization. In: SPE annual technical conference and exhibition, Denver, Colorado, USA

36. Mohaghegh S, Platon V, Ameri S (1998) Candidate selection for stimulation of gas storage wells using available data with neural networks and genetic algorithms. In: SPE eastern regional meeting, Pittsburgh, Pennsylvania

37. Nagi J, Kiong TS, Ahmed SK, Nagi F (2009) Prediction of PVT properties in crude oil systems using support vector machines. In: Proceedings of ICEE 2009 3rd international conference on energy and environment, Malacca, Malaysia

38. Nikravesh M, Aminzadeh F (2003) Soft computing for intelligent reservoir characterization and modeling. In: Soft Computing and Intelligent Data Analysis, Nikravesh M, Aminzadeh F, Zadeh LA (eds) Developments in petroleum science series, vol 51. Elsevier, Amsterdam, pp 3–32

39. Oloso MA, Khoukhi A, Abdulraheem A, Elshafei M (2009a) A genetic-optimized artificial neural networks for predicting viscosity and gas/oil ratio curves. In: SPE/EAGE reservoir characterization and simulation conference and exhibition, Abu-Dhabi, UAE

40. Oloso MA, Khoukhi A, Abdulraheem A, Elshafei M (2009b) Prediction of crude oil viscosity and gas/oil ratio curves using recent advances to neural networks. In: SPE/EAGE reservoir characterization and simulation conference, Abu-Dhabi, UAE

41. Osman EA, Al-Marhoun MA (2005) Artificial neural networks models for predicting PVT properties of oil field brines. In: 14th SPE middle east oil and gas show and conference, Bahrain

42. Osman EA, Abdel-Wahhab OA, Al-Marhoun MA (2001) Prediction of oil PVT properties using neural networks. In: SPE middle east oil show. Society of Petroleum Engineers, Manama, Bahrain

43. Saemi M, Ahmadi M, Varjani AY (2007) Design of neural networks using genetic algorithm for the permeability estimation of the reservoir. J Petrol Sci Eng 59:97–105

44. Sun Q, Li J, Zhang C, Liu X (2005) High watercut reservoir permeability prediction by flow unit, ANFIS and multi-statistics. Well Logging Technol 29(2):121–124

45. Sundgren H, Winquist F, Locker I, Lundstrom I (1991) Artificial neural networks and gas sensor arrays: quantification of individual components in a gas mixture. Meas Sci Technol 2:464–469

46. Tran QA, Zhang QL, Li X (2003) Reduce the number of support vectors by using clustering techniques. In: International conference on machine learning and cybernetics, vol 2. IEEE, pp 1245–1248

47. Varotsis N, Gaganis V, Nighswander J, Guieze P (1999) A novel non-iterative method for the prediction of the PVT behavior of reservoir fluids. In: SPE annual technical conference and exhibition, Houston, Texas, USA

48. Wang P, Pope G (2001) Proper use of equations of state for compositional reservoir simulation. J Petrol Technol 53(7):74–81

49. Wong PM, Aminzadeh F, Nikravesh M (2002) Intelligent reservoir characterization. In: Wong PM, Aminzadeh F, Nikravesh M (eds) Soft computing for reservoir characterisation and modeling, studies in fuzziness and soft computing. Springer, Berlin, pp 3–12

50. Xiao-mei L, Qin-hua S, Jian-xin L, Wei-fang L (2009) Prediction of fracture porosity of carbonate reservoir with seismic attributes, multi-analysis and ANFIS. Well Logging Technol 33(3):631–684

51. Xie D, Wilkinson D, Yu T, Ramon S (2005) Permeability estimation using a hybrid genetic programming and fuzzy/neural inference approach. In: SPE annual technical conference and exhibition, Dallas, Texas, USA

52. Zhang S, Zhang C, Yang Q (2003) Data preparation for data mining. Appl Artif Intell 17(5–6):375–381