1-1-2022

Prediction of Formation Conditions of Gas Hydrates Using Machine Learning and Genetic Programming

Anupama Kumari
Indian Institute of Technology, Roorkee, India

Mukund Madhaw

Vishnu S. Pendyala
San Jose State University, vishnu.pendyala@sjsu.edu

Recommended Citation

Anupama Kumari, Mukund Madhaw, and Vishnu S. Pendyala. "Prediction of Formation Conditions of Gas Hydrates Using Machine Learning and Genetic Programming" *Machine Learning for Societal Improvement, Modernization, and Progress* (2022): 200-224. https://doi.org/10.4018/978-1-6684-4045-2.ch010

Chapter 10
Prediction of Formation Conditions of Gas Hydrates Using Machine Learning and Genetic Programming

Anupama Kumari
https://orcid.org/0000-0003-2318-6090
Indian Institute of Technology, Roorkee, India

Mukund Madhaw
Independent Researcher, India

Vishnu S. Pendyala
https://orcid.org/0000-0001-6494-7832
San Jose State University, USA

ABSTRACT

The formation of gas hydrates in the pipelines of oil, gas, chemical, and other industries has been a significant problem for many years because the formation of gas hydrates may block the pipelines. Hence, the knowledge of the phase equilibrium conditions of gas hydrate became necessary for the economic and safe working of oil, gas, chemical industries. Various thermodynamic approaches with various mathematical techniques are available for the prediction of formation conditions of gas hydrates. In this chapter, the authors have discussed the least square support vector machine and artificial neural network models for the prediction of stability conditions of gas hydrates and the use of genetic programming (GP) and genetic algorithm (GA) to develop a generalized correlation for predicting equilibrium conditions of gas hydrates.

DOI: 10.4018/978-1-6684-4045-2.ch010

INTRODUCTION

The oil and gas industry pipelines face various flow assurance difficulties in the subsea networks of pipelines. It includes the treatment of multiple deposits of solids in the subsea flowlines from the fluid streams of hydrocarbon. These flow assurance problems may be caused by gas hydrate formation, slugging, degradation, wax, scales, etc. Out of these problems, the appearance of gas hydrates is of grave concern. A suitable environment with high pressure and low temperature in the marine climate of the oil and gas industry is the primary cause of the gas hydrate formation. Hence, deep-water flow preservation became a severe problem for these industries (Kumari, Khan, Misra, Majumder, & Arora, 2020).

Gas hydrates are solid crystalline materials formed after trapping the guest molecules inside the cage-like structure formed by water molecules. Methane (CH_4), Ethane (C_2H_6), Propane (C_3H_8), Nitrogen (N_2), Carbon Dioxide (CO_2), and Hydrogen Sulphide (H_2S) are the most available guest molecules which can form the gas hydrates (Kumari et al., 2020). These hydrates are abundant in nature. The stability of hydrates crystals is because of the Vander Waals force, which occurs because of no bonding between the water and gas molecules (Kumari, Hasan, Majumder, Arora, & Dixit, 2021). The significant deposits for the formation of in situ gas hydrates are the deep ocean regions and the permafrost zones. About 97% of gas hydrate deposits are in the form of submarine sediments (Ke & Chen, 2019). The gas hydrates are also known as clathrate. The word clathrate comes from the Greek word "Khlatron." The meaning of this Greek word is the shield.

Sir Humphrey Davy discovered the gas hydrates in 1810 (E. D. Sloan, 1998), which remain stable at temperatures higher than 273.15 K. The gas hydrate has become a significant research interest for academics and the business sector by considering many factors. Gas hydrates could become a vast energy source because the amount of energy stored in the gas hydrates is twice the total quantity of all the other forms of energy available on the earth (Chong, Yang, Babu, Linga, & Li, 2016). The gas hydrate can also become a threat to the atmosphere and environment after the accidental dissociation of methane gas from the hydrate deposits (Giustiniani, Tinivella, Jakobsson, & Rebesco, 2013). Some other applications of gas hydrates are the capture and storing of hydrate-based carbon (Cai, Pethica, Debenedetti, & Sundaresan, 2016), seawater desalination (M. S. Khan, Lal, Sabil, & Ahmed, 2019), hydrogen storage (Lee et al., 2011), distribution and storage of natural gas (Gbaruko, Igwe, Gbaruko, & Nwokeoma, 2007), and the advanced form of air conditioning systems (Fournaison, Delahaye, Chatti, & Petitet, 2004).

The formation of natural gas hydrates in the pipelines is a critical and challenging issue because of the reformation of hydrates in the production pipelines because of the higher pressure and lower temperature. The research on gas hydrates requires multiscale analysis because of the thermodynamic and phase equilibrium features of gas hydrates (Englezos, 1993).

FORMATION CONDITION OF GAS HYDRATES

Gas hydrates crystallize at high pressure and low temperature after capturing the methane or other hydrocarbon gases by the lattice structure of water molecules. The gas hydrate formation requires enough concentration of gas. The gas solubility in water and the ionic strength of water are dominant factors in forming gas hydrates (Kvenvolden, 1993). The specific temperature-pressure regime controls the stability

of gas hydrates. The significant areas available for the formation of Gas hydrates are mainly continental margins where methane-bearing fluids and sufficient organic carbon are available (Davie, Zatsepina, & Buffett, 2004). The released methane gas in the shallower zones of the earth is encaged inside the cages of the crystal lattice of water molecules at high pressure and lower temperature.

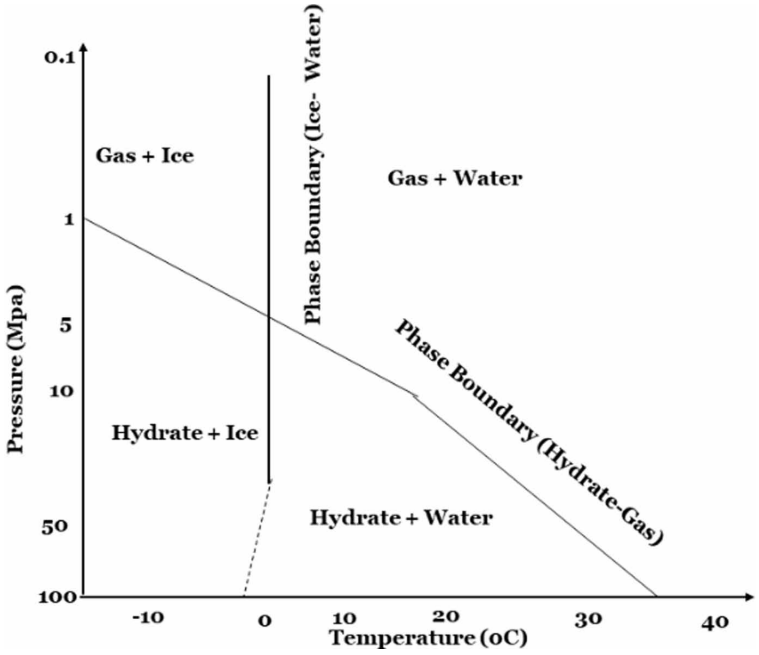
The favorable environments in the permafrost deposits are due to the low temperature and high pressure because of the lower surface and the presence of sediments. These factors prepared desired conditions for the gas hydrate formation in the earth's subsurface. At some depths in the tropical regions, the temperature at the sea bottom decreases substantially from the detected temperature at the sea's surface. The width of the water column creates desired pressure condition for the gas hydrate formation in the subsurface of the earth. There is the requirement of enough accessible methane gas having non-solubility of methane gas in the water in both environments for the gas hydrate formation. The oceans' shallower zones release methane gas because of the biochemical and microbial decay of disposing of organic materials in the bottom sediments. It may also leak because of the discharge of the natural gases from the gas-bearing deposits and shallow oil (Thakur & Rajput, 2010).

Gas hydrates are stable at in situ conditions of desired pressure-temperature values in the subsurface earth. The upper limit of the thickness of the stability zone of the gas hydrates is the seafloor and regions enclosed at the convergence of the geothermal gradient curve, and the lower limit is the equilibrium with pure and seawater. The other parameters on which the stability of gas hydrates depends are the temperature at the sea bottom, the water salinity, composition, and local geology. The accurate estimation of the width of the stability zone of the gas hydrate is difficult because the above parameters show different in situ conditions and cannot be predicted accurately (Thakur & Rajput, 2010). There are two phases (Solid-phase and vapor phase) present in the hydrate, which is in equilibrium thermodynamically. The solid and vapor phase equilibrium arises under specific pressure and temperature conditions for the desired gas. The water is present in vapor form only in natural gas and other liquid hydrocarbons. There are some different phases, such as liquid water, liquid hydrocarbon, and water ice, because of the water content of the natural gas. These additional phases are also in equilibrium with the solid-phase and vapor phases. Figure 1 shows the phase diagram for gas hydrate.

BACKGROUND

Considerable literature is available on measuring the formation condition of gas hydrates. The basic model for predicting equilibrium conditions of gas hydrates is the Vander Waal Platteeuw and Chen-Guo models (Sinehbaghizadeh, Javanmardi, Roosta, & Mohammadi, 2019). However, these models are not sufficiently accurate for the forecast. In addition to these available traditional methods, machine learning models may also be used for predicting equilibrium conditions of gas hydrates. As we describe in the following sections, the machine learning models have the potential to improve prediction accuracy. For instance, Artificial Neural Networks (ANN) and Least-Squares Support Vector Machine (LSSVM) models were recently used to predict hydrate formation conditions with significant success (Xu, Jiao, Zhang, Huffman, & Wang, 2021).

Figure 1. Phase Diagram (Collett, Lewis, & Uchida, 2000)



PREDICTING GAS HYDRATES FORMATION CONDITIONS

Natural gas hydrate formation is like a thermodynamical operation that connects temperature, pressure, and gas composition. Thermodynamic rules can predict the pressure or temperature for hydrate formation. The temperature and pressure are the main parameters for analyzing hydrate stability conditions. Hydrate will form at low pressure if the quantity of heavy components in natural gas is enhanced. Many datasets are available in the literature for gas hydrates' formation conditions, which can help predict the formation conditions of gas hydrates through Genetic Programming (E. Dendy Sloan & Koh, 1998).

Genetic programming can develop different nonlinear correlation models using pressure-temperature data, forming the gas hydrates. Each data set includes two groups randomly: training and test sets. The training data set is used to produce the prediction model, and the test set checks the prediction efficiency of the developed model. This tool performs the multi-gene symbolic regression analysis by generating the central equation and then crossing over the best equation and producing the best equation to estimate the formation conditions of gas hydrates. The formation temperature, pressure data, and molecular weight are the inputs for the genetic programming, and then the subset correlation can be obtained with the excellent value of evaluation parameters (Abooali & Khamehchi, 2017).

Other than conventional thermodynamic models, various statistical and mathematical algorithms are available to predict the hydrate stability conditions. Genetic programming (Abooali & Khamehchi, 2017), Artificial Neural Networks (ANN), and generalized regression neural networks are well-known tools (Zahedi, Karami, & Yaghoobi, 2009) for estimating hydrate formation conditions. But genetic programming is one of the most effective methods for predicting and optimizing processes with sound capability. Genetic programming can easily predict the formation conditions of gas hydrates without any

nested terms. The nested terms can give complicated background calculations and intensive calculations (S. H. Khan, Kumari, Chandrajit, & Amit, 2020).

DATASET AND PRE-PROCESSING

Abundant research literature is available on predicting gas hydrates' formation conditions by genetic programming. Abooali and Khamehchi developed three new data-based models for predicting the formation temperature of natural gas hydrate as a function of equilibrium pressure and molecular weight of gas using genetic programming with an acceptable error percentage (Abooali & Khamehchi, 2017). Khan et al. 2020 developed a new correlation by genetic programming to predict the formation pressure of hydrates with a reasonable error percentage (S. H. Khan et al., 2020). Cao et al. 2020 combined the genetic algorithm and support vector machine to develop a new model for estimating gas hydrate formation conditions with a slight average relative deviation (Cao, Zhu, Li, & Han, 2020).

The genetic algorithm's first step for predicting the stability conditions of gas hydrates is guessing the random correlation or function of pressure, temperature, and specific gravity of gas and the initial values of constants. The fitness of the process depends upon the calculated value of the selected function for each constant. A more suitable value of constants with the low values of the fitness function is determined and then allowed to perform the next step. The columns selected for the dataset are pressure, temperature, and specific gravity or molecular weight of gases. In machine learning, dimensionality is the number of features of the data set. Feature selection helps remove irrelevant dimensions from the selected data sets. Dimensionality reduction slashes the number of random variables or attributes down while preserving the significant characteristics of the data. In a genetic algorithm, dimensionality reduction is achieved by the mutation and crossover of the datasets. The normalization technique calculates the optimum values of the constants for the manipulation of data. Data manipulation is performed by increasing or decreasing the range of data before it goes to the next step.

The best method for predicting phase equilibria conditions of gas hydrates is by experimentation. The experimental observation of gas hydrates' pressure and temperature conditions for various gas compositions is not usually feasible. Hence, the calculation of equilibrium conditions of gas hydrates by using correlation techniques is needed.

Two experimental techniques are available to study the phase equilibria of gas hydrates. The first is a dynamic method, where a given gas is injected inside a chamber continuously at favorable gas hydrate formation conditions, usually at low temperatures. The formation of gas hydrates stops the continuous flow of gas from the enclosure at the desired pressures to equilibrate the system to form gas hydrates. After the formation of gas hydrates, the pressure or temperature of the system can be decreased or increased, respectively (Kumari, Madhaw, Majumder, & Arora, 2021). The second is a static method, which shows the growth of gas hydrate crystals in an autoclave or stationary high-pressure vessel. In this method, the study uses the controlled temperature and pressure of the cell. This method applies to studying the phase equilibria of gas hydrates because of the ease of measuring intensive properties (Kumari, Madhaw, Majumder, & Arora, 2021).

This chapter discusses the applicability of LSSVM, ANN, and genetic programming as essential mathematical tools for predicting hydrate formation conditions. These tools are particularly helpful from a big data point of view when the dataset has wide ranges of pressure, temperature, and molecular

weight of gases. GA, ANN, and LSSVM have been used to model the relationship between the growth rate of methane hydrate, pressure, and temperature.

ARTIFICIAL NEURAL NETWORKS FOR GAS HYDRATES PREDICTION

Artificial neural networks (ANNs) use various computational units called neurons parallelly connected in an enormous structure. They do not require an explicit conceptualization of the physical or mathematical relationships. These neural networks are established to be universal functional approximators, that is, they can compute any function f(x) approximately, where x represents the inputs. The neural networks are used as a computational algorithm to learn, organize, generalize, and cluster data (Blackwell & Chen, 2009). The frequently used neural networks are the feed-forward neural networks (FNN) which contain one input and output layer and a number of hidden layers in between (Schmitz, Zemp, & Mendes, 2006). The FNNs are flexible, simple, and easy to use, and these networks can execute differentiable and continuous mathematical functions, which can minimize the error analysis (Zahedi et al., 2009). The ideal number of neurons in the hidden layer is a hyperparameter of the FNNs that needs to be determined. Few neurons generate a network with low accuracy, and a higher number can lead to overfitting and then a poor standard of extrapolation and interpolation. Methods such as Bayesian regularization with a Levenberg-Marquardt algorithm apply with the FNNs to overcome the disadvantage (Marquardt, 1963).

The neural network's input layer accepts normalized input data and forwards it to the hidden layers of the network. The network processes these data from the input layer with the help of weighted interconnections between the neurons in the layers. In a fully connected network, every i^{th} neuron in a k^{th} layer can be attached to every neuron in the adjoining layers. The i^{th} neuron in the k^{th} hidden layer executes the summation of all weighted inputs and propagates the result, typically through a nonlinear activation function f. This process is expressed as equation (1). The propagations continue in this fashion on the adjoining neurons of the next hidden layer and eventually, the output neurons. Like in any machine learning task, the data sets are subdivided into training, validation, and test data sets. The training data set is used to determine the network's parameters. The validation data set is used to fine-tune the model's performance. After the model gets trained, the test data is used to evaluate the performance of the trained model. All the biases and synaptic weights are initialized randomly. During the process of training the network, an optimization algorithm adjusts the synaptic weights until the error in the predicted output is minimized (Chapoy, Mohammadi, & Richon, 2007).

$$O_j = f(\sum w_{ji} x_i + b_j) \tag{1}$$

where w_{ji} is the weight, x_i is the input value, b_j is the threshold value, and f is the activation function.

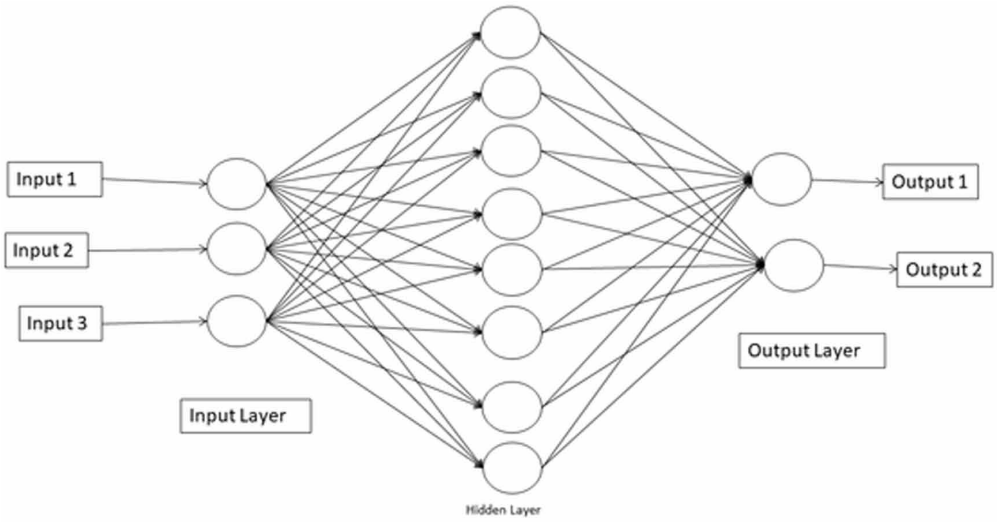
The output from one neuron becomes the input to the neuron in the next layer. This simple arrangement of the interconnected neurons arranged in layers is powerful enough to approximate any mathematical function of the inputs. The layers of neurons and their interconnections using linear and non-linear transformations can convert the input data to any desired output. For instance, input pixel values of images can be converted to probabilities of the image being that of a lion or a tiger. For this, the weights of the interconnections obviously need to change to suit the desired output. The weights of the neural network are adjusted by an optimization algorithm, for example, stochastic gradient descent using backpropagation, in order to minimize the error between the generated and actual values of the target. A common

function to compute the error often used in machine learning regression problems when the output values are continuous is the mean square error, which is the average value of the squared difference between the network's output and the true output in the training dataset.

$$Error = \frac{1}{N}\sum_{i=1}^{N}\left(P_i - C_i\right)^2 \tag{2}$$

In the above equation, N is the number of training data, P_i and C_i are the targeted and calculated output, respectively (Ghavipour, Ghavipour, Chitsazan, Najibi, & Ghidary, 2013). Figure 2 shows a schematic of a three-layered neural network.

Figure 2. A three-layered neural network (Kumari, Madhaw, Majumder, & Arora, 2021)



Elgibaly and Elkamel 1998 calculated the formation pressure of the gas hydrate with an average error of 19%. They also evaluated the amount of thermodynamic inhibitors required in a different system (Elgibaly & Elkamel, 1998). Chapoy et al. 2007 used ANN to predict formation conditions of natural gas hydrate in the absence and presence of inhibitors (Chapoy et al., 2007). Moradi et al. 2013 fitted an ANN model to the experimental data of the gas hydrates of binary hydrocarbon mixtures (Moradi, Nazari, Alavi, & Mohaddesi, 2013). Mohammadi and Richon 2010 and Mohammadi, Belandria, and Richon developed ANN models to predict hydrogen and methane hydrate formation conditions in the presence of the promoters (tert-butylamine and tetrahydrofuran). The results of the ANN model are not well fitted with the results of the thermodynamic models (Mohammadi, Belandria, et al., 2010; Mohammadi & Richon, 2009).

Babakhani et al. 2015 observed a relative error of 1.02% after evaluating an ANN model (Babakhani, Bahmani, Shariati, Badr, & Balouchi, 2015). Mohammadi et al. 2010 calculated the formation conditions of tetrahydrofuran and methane, nitrogen or carbon dioxide hydrate, hydrogen, and tetra-n-butyl ammonium bromide hydrate. The developed ANN model evaluates the acid gas systems and obtained

satisfactory results (Babakhani et al., 2015; Ghavipour et al., 2013; Mohammadi, Martínez-López, & Richon, 2010; ZareNezhad & Aminian, 2012). Rebai et al. 2019 developed an ANN model to calculate the thermodynamic properties depending on the formation of gas hydrates for more mixtures. The developed model also evaluated some thermodynamic properties of gas hydrates, such as viscosity (Rai, Majumdar, DasGupta, & De, 2005; Rebai, Hadjadj, Benmounah, Berrouk, & Boualleg, 2019), density, compressibility factor (Bouchard & Granjean, 1995), vapor pressure (Laugier & Richon, 2003), heat transfer coefficient (Potukuchi & Wexler, 1997) and vapor-liquid equilibrium (Ganguly, 2003; Petersen, Fredenslund, & Rasmussen, 1994; Sablani, Baik, & Marcotte, 2002; Sharma, Singhal, Ghosh, & Dwivedi, 1999).

Heydari et al. 2006 and Zahedi et al. 2009 trained an ANN model to calculate the hydrate formation temperature (Heydari, Shayesteh, & Kamalzadeh, 2006) (Zahedi et al., 2009). Mehrizadeh used ANNs to predict the formation conditions of gas hydrate in the presence of thermodynamic inhibitors (Mehrizadeh, 2020). Soroush et al. 2015 used the ANN model to predict the formation temperature of the natural gas hydrate. They observed a total mean square error of 0.349 only (Soroush et al., 2015). Mesbah et al. 2017 developed a supervised learning algorithm, namely, MLP ANN, to predict the dissociation pressure of the semi-clathrate hydrate for carbon dioxide and its binary mixtures. They observed the AARD% of 3.13 and overall R^2 as 0.9961(Mesbah, Soroush, & Roham, 2017).

Several datasets for the stability conditions of gas hydrates were used to develop an ANN. The main parameters for applying ANN to determine stability conditions of gas hydrates are temperature, pressure, and the specific gravity of the gas. The training and testing set apply to the experimental data points of gas hydrates. The input layer of the ANN takes features like the pressure and specific gravity of gases. The number of hidden layers is a hyperparameter and depends on the data set. If the input parameters for developing an artificial neural network are pressure and specific gravity of gases, and then the temperature will be the output parameter. The hyperparameters of an artificial neural network are the variables that determine the topology of the neural network and the variables that determine the neural network's training step. The number of hidden units is a hyperparameter too, and these hyperparameters are set before the training step and before optimizing weights and bias terms. The hidden layers are the layers between the input and output layers, and the optimum number of hidden layers depends upon the minimum error value.

The following steps are involved in setting up an ANN model for prediction.

1. Input data comprising temperature and specific gravity of gases is selected based on the experiments. This is the ground truth from which the training, validation, and test datasets are obtained.
2. The target variable is pressure.
3. The input data become the input for the neural network, and the target variable becomes the target for the output of the ANN.
4. The topology of the neural network model depends on the number of layers and the number of neurons in each layer. These "hyperparameters" are chosen accordingly.
5. Levenberg-Marquardt Backpropagation is commonly used for training the network.
6. The predicted outputs from ANN are compared with the expected predictions from the training data and loss calculated.
7. The training process progresses in iterations until the loss computed based on the training data is minimized and the network parameters are determined.
8. The selected parameters apply to the neural network for the dataset's training, testing, and validation.

9. The above procedure repeats for multiple configurations to determine the best hyperparameters.

THE LEAST-SQUARES SUPPORT VECTOR MACHINE

Vapnik's support vector machine (SVM) builds upon the statistical learning theory (V. N. Vapnik, 1999). SVM is a supervised machine learning method, and LSSVM is the reformulation of SVM. LSSVM uses the least-squares objective function, thereby circumventing quadratic programming (Yarveicy & Ghiasi, 2017a). SVM is a skilful learning method adopted for regression analysis and classification in machine learning. Still, the main drawback of the SVM method is its complexity in the computational procedure (V. Vapnik, 1999). This mathematical tool can solve several complex problems, from the nonlinear function approximation to pattern classification. This algorithm obtains an optimum hyperplane with the minimum distance from the actual data by projecting the input features to a higher dimension in case the data is not linearly separable (Soroush, Mesbah, Shokrollahi, Bahadori, & Ghazanfari, 2014). In this method, the chances for overfitting are less, and there is no need for multiple adjustable parameters (Cortes & Vapnik, 1995). LSSVM is generally preferred over SVM for scalability reasons.

The solution of quadrating encoding equations increases the complications in the optimization procedure. The LSSVM reduces the complexity of SVM and then enhances the speed. Then LSSVM model performs the linear calculations by linear encoding. The network results after resolving given fixed linear equalities in the LSSVM algorithm (Suykens and Vandewalle, 1999). The LSSVM can operate on large data sets with acceptable accuracy by changing the inequality constraints with the equality constraints (Pelckmans et al., 2002). The regression error adapts as an additional constraint in the LSSVM algorithm. This error calculates by the difference between the predicted and experimental data (Pelckmans et al., 2002).

Suppose a data set with the given equation (3) in which x and y are the input data points. (For the gas hydrates, x and y should be temperature and specific gravity of gases.)

$$D = \{(x_1,y_1) \ldots (x_k,y_k) \ldots (x_N,y_N), x_k \in R^n, y_k \in R\} \tag{3}$$

Equation (3) is generalized by the following equation of nonlinear function (4)

$$f(x) = wg(x) + b \tag{4}$$

Equation $w \in R^{nh}$ is the weight vector in the initial weight space, b is the bias term, $g(\circ)$: $R^n \rightarrow R^{nh}$ is the nonlinear mapping, and n_h is the space dimension. This optimization problem is given by

$$\min J\left(w,e\right) = \frac{1}{2} w^T w + \frac{1}{2} \mu \sum_{k=1}^{n} e_k^2 \tag{5}$$

μ is the relative weight of the summation of the errors calculated during regression after comparing the weight of the regression.

The above equation (5) gives the constraints of equation (6)

$$y_k = (w^T g(x)) + b + e_k,\ k=1,2,...,n \tag{6}$$

In equation (6), e_k is the error, w^T is the weight matrix transpose, and $\mu \geq 0$ is the regularization constant. After solving (equation (6)) in the (equation (5))

$$L_{LSSVM} = \frac{1}{2} w^T w + \frac{1}{2} \mu \sum_{k=1}^{n} e_k^2 - \sum_{k=1}^{n} \beta_k \left\{ \left(w^T g\left(x\right) \right) + b + e_k - y_k \right\} \tag{7}$$

The Lagrangian multipliers $\beta_k \in R$. The conditions needed for the optimization are given in equation (8) to equation (11).

$$\frac{\partial L_{LSSVM}}{\partial b} = 0 \rightarrow \sum_{K=1}^{n} \beta_k = 0 \tag{8}$$

$$\frac{\partial L_{LSSVM}}{\partial w} = 0 \rightarrow w = \sum_{K=1}^{n} \beta_k g(x_k) \tag{9}$$

$$\frac{\partial L_{LSSVM}}{\partial \beta_k} = 0 \rightarrow \left(w^T g\left(x\right) \right) + b + e_k - y_k = 0 \ldots\ldots, \left(k = 1, \ldots\ldots n \right) \tag{10}$$

$$\frac{\partial L_{LSSVM}}{\partial e_k} = 0 \rightarrow \beta_k = \mu e_k, \quad \left(k = 1, \ldots\ldots n \right) \tag{11}$$

Equation (12) will obtain after assuming a linear regression relationship for the independent and dependent variables.

$$\mathrm{y} = \sum_{K=1}^{n} \beta_k x_k^T x + b \tag{12}$$

We use a kernel function for the dot product in equation (12) to enhance the applicability.

$$\mathrm{y} = \sum_{K=1}^{n} \beta_k K\left(x_k, x\right) + b \tag{13}$$

The Kernel function of $K(x_i,x)$ is the dot product of vectors $g(x_i)$ and g(x) in the higher dimensional space (equation (14))

$$K(x,x_k) = g(x) g(x_k)^T \tag{14}$$

The RBF kernel given in equation (15) and the polynomial kernel given in equation (16) are popularly used in the LSSVM model.

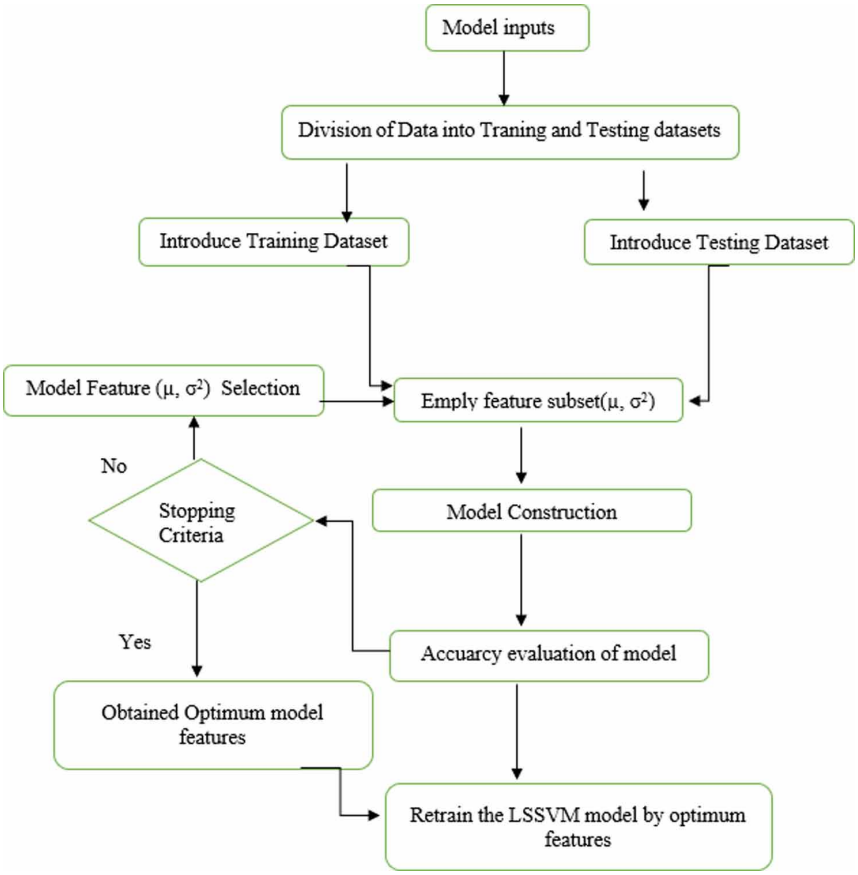
$$K(x,x_k) = \exp(x_k - x^2 / \sigma^2) \tag{15}$$

$$K(x,x_k) = (1 + x^T x_k / c)^d \tag{16}$$

The σ^2 is the squared width. It is a hyperparameter. In the polynomial kernel, d is the degree, which is also a hyperparameter.

Figure 3 shows the steps involved in developing the LSSVM model to predict the stability conditions of gas hydrates. The algorithm starts with the empty feature set, and then after each iteration, this adds the feature. The regularization (μ) and the kernel (σ^2) parameters should specify before the training step.

Figure 3. Schematic diagram of the LSSVM model (Kumari, Madhaw, Majumder, Arora, & Dixit, 2021)



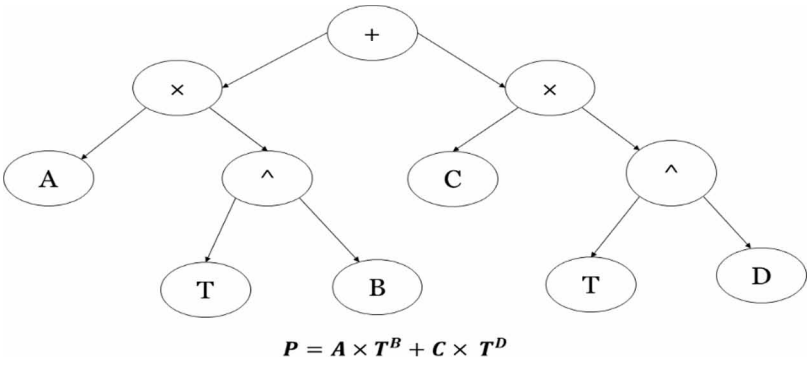
Mesbah et al. 2017 used the LSSVM model based on the gas hydrate structure and provided an effective algorithm with a squared correlation coefficient of 0.9918 (Mesbah, Soroush, & Rezakazemi, 2017). Baghban et al. 2016 applied the LSSVM model to the Katz Chart data points and predicted the

formation temperature of gas hydrates with a squared correlation coefficient of 0.9973 and minimum square error of 0.778634 (Baghban et al., 2016). Yarveicy and Ghiasi 2017 utilized a model that uses extremely randomized trees to predict the formation conditions of gas hydrate of different gases formed in pure water, salts, and alcohols. This model provides a squared correlation coefficient of greater than 96 and a percentage absolute relative deviation (%ARD) between 0.04 and 0.32 (Yarveicy & Ghiasi, 2017b). Mohammadi et al. applied the LSSVM model to estimate phase equilibria for the CO_2, CH_4, or C_2H_6 hydrate in porous media (Mohammadi et al., 2012).

GENETIC ALGORITHM

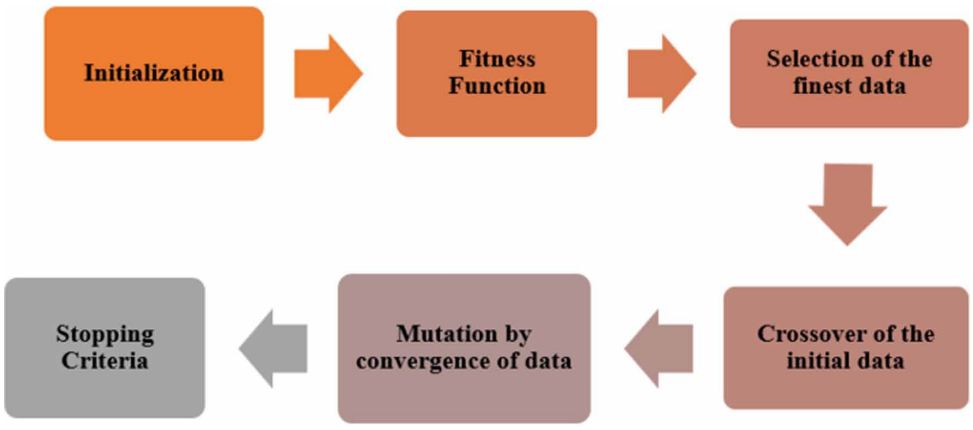
The Genetic Algorithm (GA) was developed in 1975 by John Holland, supported by the concept of the artificial system (Holland, 1992). This algorithm is appropriate for searching and optimization problems. GA performs various operations, and every procedure carries multiple techniques, which give different results due to the diverse nature of methods (Abbasi & Hashim, 2016). An off-shoot of GA called Genetic programming (GP) is a robust methodology whose programming procedure depends upon biological generation systems. It is an efficient technique applied in different optimization and mathematical problems. Genetic Programming (GP) is a popular technique to determine correlations in modeling projects. Genetic Programming (GP) solves a problem by generating a population randomly using mathematical equations like genes. These genes are chromosome-like structures of syntactic trees from the input data. GP develops several genes usually instead of one to enhance the accuracy of the prediction. A developed multi-gene network contains one or more genes and this provides more specific functions frequently than the additional developed models with one monolithic GP tree. Figure 4 shows an example of a simple tree structure.

Figure 4. An Example of the Tree structure for Genetic Programming (S. H. Khan et al., 2020)



The genetic algorithm endeavors to obtain the optimal solution to a given problem. The algorithm is quite widely used in economics, bioinformatics, and data sciences. Genetic Algorithms employ ideas from the theory of human genetics. These algorithms provide a heuristic approach to solving complex problems, and these can also be used to optimize the results. The Genetic Algorithm depends upon Initialization, Selection, Crossover, Mutation, and Termination (Figure 5).

Figure 5. Steps involved in Genetic Algorithm



Initial Population

The algorithm starts with initializing data sets (population), and each data is a solution to the given problem. A group of parameters or variables characterizes each piece of information.

Fitness Function

A fitness function uses to check the fitness of data. This fitness function identifies the ability of a data to compete with the other data. The probability of data identification for reproduction depends on its fitness value.

Selection

The fittest data sets are selected and then sent for the following data generation in the selection step. Data with a high fitness value have more chances to be chosen for reproduction. Two pairs of data are selected based on their fitness value for the next step.

Crossover

The most significant step in the genetic algorithm is crossover. A crossover point is selected randomly for each data set to be mated.

Mutation

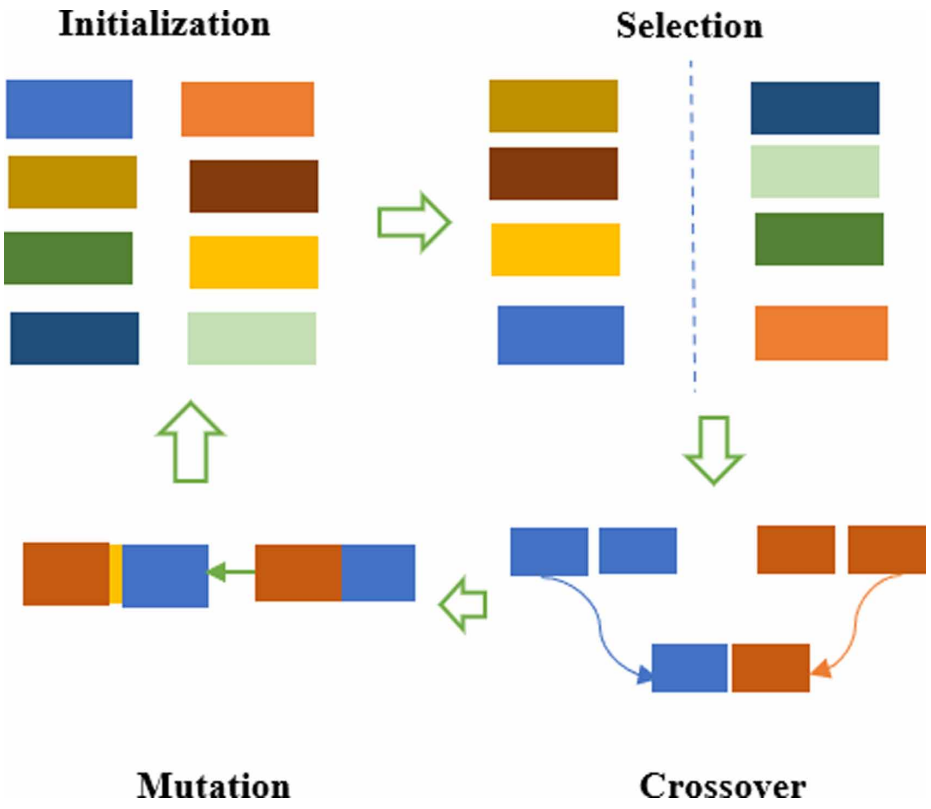
Some data sets send for the mutation with a low random probability. The mutation maintains diversity within the data sets and avoids premature convergence of data.

Termination

The genetic algorithm stops when the data converge and then provides solutions for the given problem.

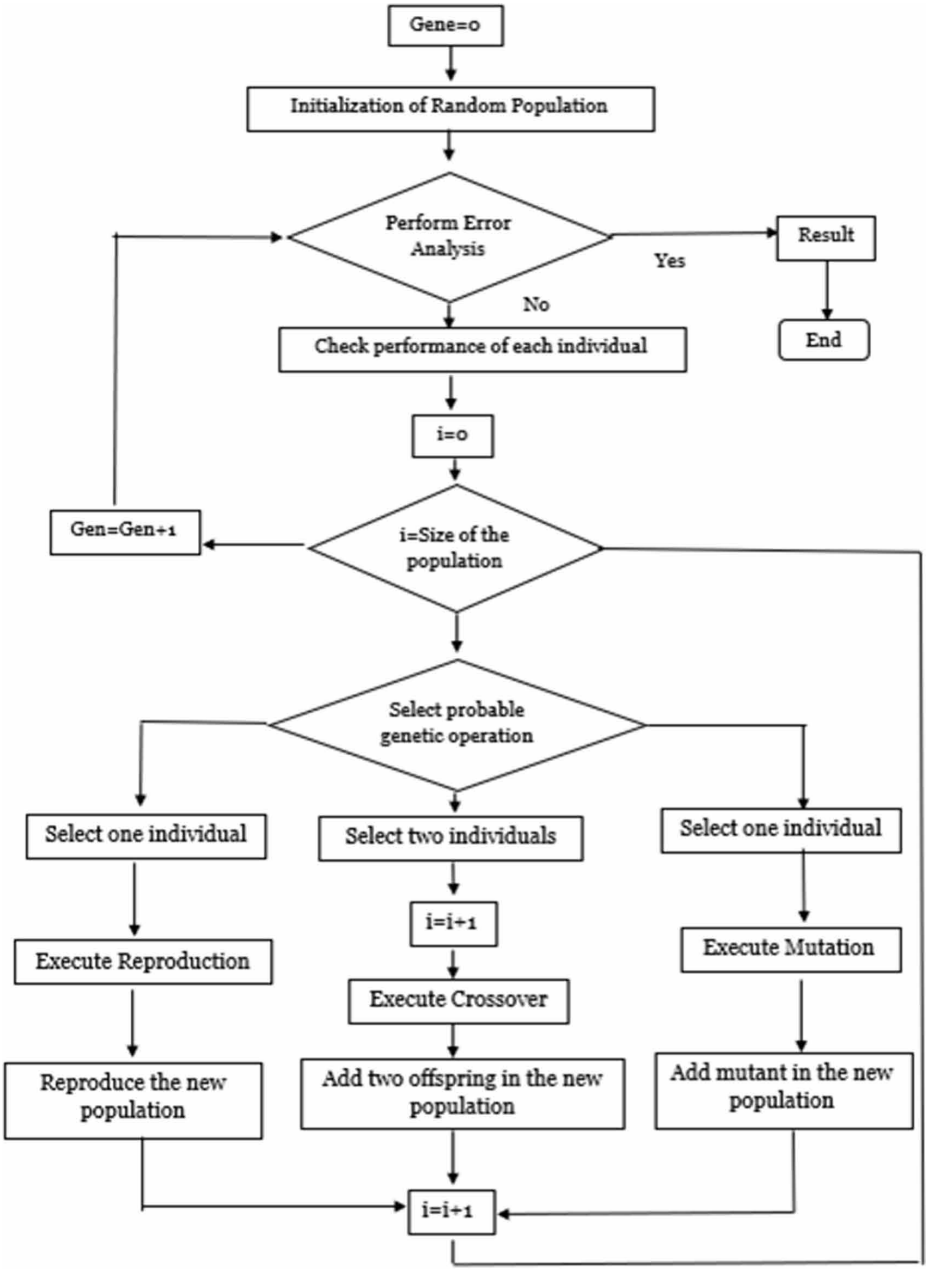
Hence, a genetic algorithm is an optimization technique applied to find input data sets to obtain the best output. Figure 6 shows a schematic representation of all steps involved in the genetic algorithm.

Figure 6. Representation of steps for Genetic Algorithm



GP technique is a symbolic regression method. The algorithm itself explores the model and fits the constants afterward. In ordinary regressions, the structure of the model must be stated by the user and subsequently fitted to the model constants. The inclusive status of the primary model is attained by the weighted summation of every available function after the generation of the first population randomly. The genes with a biased term constitute these functions: the bias and weight terms obtained by the standard least square method. The first accepted population may not provide a good prediction; the structure often requires updating, and the unusable functions change with others. Hence, the algorithm goes through the best executing trees and then improves the network to obtain a new population. New structures create the genes in the new population and repeat the summation of the weights of all newly developed genes. Then the new bias and weight term are predicted. The regeneration process performs repeatedly, and then the mathematical functions are also iterated to obtain new populations. This iteration will stop after solving the problem successfully (Jain, 2017). Figure 7 shows the overall flowchart of GP.

Figure 7. Overall Flowchart for Genetic Programming



Searson developed a free, open-source tool to perform genetic programming in MATLAB software. This tool creates the nonlinear correlation to determine the temperature for the hydrate formation. A toolbox "GPTIPS" develops for multi-gene symbolic regression problems. All the required steps of GP perform to obtain the best correlation. For each subset, input data, including experimental sets of hydrate formation Pressure, temperature, and Molecular weight, were provided to the GP program. Correlation for this subset acquires the desired accuracy.

Abooali and Khamehchi 2019 applied three new data-based models using genetic programming to predict the formation temperature of natural gas hydrates as a function of the molecular weight of gas and equilibrium pressure of gas hydrates. They observed the correlation coefficient value as 09673 for the 891 experimental data items. Abbasi and Hashim 2015 developed a model using genetic programming to predict the formation conditions of gas hydrates in subsea pipelines in the presence and absence of inhibitors (Abbasi & Hashim, 2015). Khan et al. 2020 developed a correlation for predicting the formation pressure of gas hydrates with the RMSE value of 0.68 (S. H. Khan et al., 2020).

EVALUATION CRITERIA

The perfection of the newly obtained model obtain by some familiar statistical measures like average absolute relative deviation percent (AARD%), squared correlation coefficient (R^2) (Cao et al., 2020), root means square deviation (RMSD) (Abooali & Khamehchi, 2017) and absolute average deviation (AAD%) (Seif & Kamran-pirzaman, 2020). The Low values of AARD% and RMSD and high value of R^2 prefers for evaluating the model.

$$\mathrm{AARD}(\%) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{Experimental - Predicted}{Experimental}\right| \tag{17}$$

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}\left|Experimental - Predicted\right|^2}{N}} \tag{18}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left|Experimental - Predicted\right|^2}{\sum_{i=1}^{N}\left|Experimental - Predicted\right|} \tag{19}$$

$$\mathrm{AAD} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Experimental - Predicted}{Experimental}\right| \tag{20}$$

N indicates the number of all data points (Kumari et al., 2020).

APPLICATION OF GENETIC PROGRAMMING FOR GAS HYDRATES

Another technique for predicting the stability condition of gas hydrates uses a correlation model developed through genetic programming. GP randomly generates a population of different programs shown by trees. Each tree set generates a new population using mutation, crossing over with the best-performing trees to obtain a new data set (Koza, 1994). To determine the equilibrium condition of gas hydrates by GP, pressure or temperature and specific gravity or molecular weight of gas are the input variables, and

temperature or pressure are output variables. Several steps perform to obtain a correlation, such as the primary population of the program, fitness evaluation of each program, development of the following population by mutation, crossover, and reproduction. These steps repeat many times until the termination criteria are met. The following are the important steps to achieving the GP.

1. Generate the initial population randomly from a predefined set of functions as f= (-, /, sin, cos etc) and termination function, t= (T, constant).
2. Set the size of the initial population.
3. Evaluate the programs by ramped half and half methods with some depth value. It creates disturbed and small trees.
4. Perform the error analysis of each program using the error evaluation criteria.
5. Perform the reproduction of the next generation in a probabilistic manner.
6. Perform the crossover events using the standard GP crossover method by randomly choosing a node as the crossover point.
7. Perform the mutation by applying the standard GP mutation by randomly selecting a particular program's mutation point and constant probability distribution.
8. Remove and replace another tree from the mutation point.
9. Repeat 3 to 8 steps until the termination criteria are satisfied.

Several correlations develop for the stability conditions of gas hydrates using genetic programming (Table 1). These correlations obtain by performing the genetic algorithm on the data sets.

Various researchers applied the Genetic Algorithm with the Artificial Neural Network and LSSVM model to find a new correlation for predicting stability conditions of gas hydrates. The combination of Genetic Algorithms can enhance the capability of ANN and LSSVM models. Various points consider during the development of GA-ANN and GA-SVM models, such as diversity in data points, selection of variables, and optimizing parameters. Modeling gas hydrate formation conditions can be improved by considering the operational requirements and various optimization algorithms. This adaption can give the method better accuracy than the available methods.

DISCUSSION

The above-discussed algorithms have some advantages and disadvantages for determining the stability conditions of gas hydrates. The LSSVM is considered the most powerful and reliable algorithm. The following points are some benefits of LSSVM over ANN.

1. The LSSVM model has better chances of convergence to the optimum value.
2. LSSVM is a standard algorithm that obtains a relatively quick estimation of the solution.
3. There is no requirement to know the network topology in advance like for ANN.
4. LSSVM is less prone to overfitting.
5. There is no requirement to select the number of hidden nodes.
6. LSSVM has acceptable generalization performance.
7. LSSVM has a smaller number of adjustable hyperparameters.
8. The LSSVM model follows convex optimization procedures. (Mohammadi et al., 2012)

Table 1. Available correlations developed with the genetic algorithm for the prediction of the stability condition of gas hydrates

S.No.	Correlation	References
1	1. $16.04 \le M_w \le 28.966$ (Except methane and propane) $T = 10.38\ln\left(\left(P+6.711388\right)M_w\right)+3.507\ln\left\lvert M_w-2\ln\left(M_w\right)-12.01\right\rvert$ $-34.11\ln\left\lvert\ln\left(M_w-13.45\right)\right\rvert+0.7009\ln\left\lvert M_w-\ln\left(M_w\right)-13.932642\right\rvert$ (21) $+10.58\ln\left\lvert M_w-\ln\left(M_w^2+3.664M_w\right)-12.01\right\rvert-35.34$ 2. $28.966 \le M_w \le 58.1203$ (Except methane and propane) $T = 691.6\ln\left(P+\ln\left(P^2+2P\right)\right)+0.0003531(P^2-\ln\lvert M_w$ $-P-\ln\left(P+M_w\right)\rvert)-0.08179Pln\left(P\right)+258\ln\lvert {M_w}^2-3M_w$ (22) $+P\left\lvert-645.6\ln\left(P\left(M_w+2\right)+\ln\left\lvert M_w-P\right\rvert\right)+7.063M_w\ln\left\lvert\ln\left\lvert\ln\left(2P\right)\right\rvert\right\rvert+170.5$ 3. For methane and propane $T = 3405\ln\lvert\ln(M_w+\ln\left\lvert P-8.948\right\rvert)$ $\left\lvert-800.6\ln\left(\left(M_w+\ln\left(P+4.637\right)\right)\times\ln\left(M_w-9.399\right)\right)+126.6\ln\right\rvert$ (23) $11.65143+\ln\left(M_w\right)-M_w\rvert-878.5$ The temperature is in ^{0}F, and the pressure is in psia. The molecular weight (Mw) ranges obtain by performing the genetic algorithm. These data depend on the minimum value of error obtained from each set. In these equations, the | represents the absolute operator.	(Abooali & Khamehchi, 2017)
2	For γ ranges from 0.656 to 0.787, and a pressure range from 0.58 to 8.68 MPa $\frac{1000}{T} = 4.343295+1.0734\times10^{-3}\times P-9.1984\times10^{-2}\times\ln\left(P\right)-1.071989\times\gamma_g$ (24) The temperature in K and Pressure in MPa.	(Sun, Chen, Lin, & Guo, 2003)
3	$T_K = 242\times{\gamma_g}^{0.2}\times P_{kPa}^{0.021}$ (25) The temperature in K and Pressure in kPa.	(Chavoshi, Safamirzaei, & Pajoum Shariati, 2018)
4	$T_{0_F} = -20.928+13.623\left(lnP_{psia}\right)+29.67\left(ln\gamma_g\right)-0.006\left(P_{psia}\right)\left(\gamma_g\right)^2+4.14$ $\times10^{-6}({P_{psia}}^2\left(\gamma_g\right)^3+0.979\left(lnP_{psia}\right)\left(ln\gamma_g\right)+0.19\left(P_{psia}\right)\left(\gamma_g\right)+1.25$ $\times10^{-20}({P_{psia}}^6\left(\gamma_g\right)^8+0.001(lnP_{psia})^5+281.743\left(\gamma_g\right)^7+1.25$ (26) $\times10^{-27}({P_{psia}}^8\left(\gamma_g\right)^8+7.24\times10^{-28}({P_{psia}}^8\left(\gamma_g\right)^7+0.002(lnP_{psia})^6\left(ln\gamma_g\right)^8$ $+1.84\times10^{-5}(lnP_{psia})^7\left(ln\gamma_g\right)^4+0.792\left(lnP_{psia}\right)\left(ln\gamma_g\right)^3$ The temperature in ^{0}F, and the pressure in psia.	(Khamehchi, Shamohammadi, & Yousefi, 2013)

continues on following page

Table 1. Continued

S.No.	Correlation	References
5	$P = 83904.22 \times T^{1.788} + 23529.51 \times T^{2.136} - 59133.632 \times T^{2.0435} + 2.0435 \times e^{\frac{-1218.998}{T}} - 74357563.76 \times e^{\frac{-1331.419}{T^{(1-2)*}}}$ (27) The temperature in K and Pressure in MPa.	(S. H. Khan et al., 2020)
Where M_w – Molecular weight of the gas T- Temperature P- Pressure γg_ *Specific Gravity of Gas*		

The genetic algorithm can be developed alone or in combination with LSSVM and ANN for better performance of the models. The genetic algorithm can be used as an optimization process to minimize the error obtained in applying the LSSVM and ANN model.

EXPERIMENTAL DATA

The experimental data points for predicting the phase equilibrium of gas hydrate used for the modeling are from the Katz chart (D. L. Katz, 1945; D. L. V. Katz, 1959) and Sloan (E. Dendy Sloan & Koh, 1998). The temperature and pressure range for the modeling was 273.7 K to 303.1 K and 0.6 MPa to 62.85 MPa, respectively, and the range of gravities of gas was between 0.57 to 0.86.

The radial basis function (RBF) was applied as the kernel function to develop the LSSVM model. The results of the LSSVM model are further optimized by the added genetic algorithm in the LSSVM model.

The following table 2 shows the error obtained in developing LSSVM, ANN, and GA algorithms to predict stability conditions of methane gas hydrates. Several data points are used from the literature to see the applicability of these algorithms. The error obtained after applying the LSSVM model is minimal compared to the ANN and Genetic Algorithm.

Table 2. AAD (%) obtained by different methods

	(Abooali & Khamehchi, 2017)	(Sun et al., 2003)	(Chavoshi et al., 2018)	(Khamehchi et al., 2013)	(S. H. Khan et al., 2020)	ANN	LSSVM
AAD (%)	0.9439	0.9813	0.8791	0.9609	0.32	0.9782	0.2134

CONCLUSION

This chapter discussed different models for predicting formation conditions of natural gas hydrates dependent upon the pressure and molecular weight of the gas. A compelling new model created by genetic programming seems to be one of the most effective mathematical-statistical methods for this problem

domain. The wide ranges of pressure, temperature, and molecular weight can increase the prediction ability of the developed models. Applying genetic programming to estimate phase equilibrium conditions can help design an economical technology for the dissociation of gas hydrates. Exploring gas from gas hydrates can fulfill the world's energy requirement for many centuries. This technique applies to the gas hydrate system containing salts, inhibitors, and promoters. There is scope for expansion of the application of genetic programming for the gas hydrate system with inhibitors, promoters, and many more. The solutions discussed in this chapter can help alleviate the problem of hydrate formation in pipelines of oil and gas industries.

REFERENCES

Abbasi, A., & Hashim, F. M. (2015). Prediction of hydrate formation conditions in the subsea pipeline with genetic algorithm. *Proceedings of the 2015 International Conference on Technology, Informatics, Management, Engineering and Environment, TIME-E 2015*, 133–136. 10.1109/TIME-E.2015.7389761

Abbasi, A., & Hashim, F. M. (2016). Prediction of hydrate formation conditions in subsea pipeline with genetic algorithm. *Proceedings of the 2015 International Conference on Technology, Informatics, Management, Engineering and Environment, TIME-E 2015*, 133–136.

Abooali, D., & Khamehchi, E. (2017). New predictive method for estimation of natural gas hydrate formation temperature using genetic programming. *Neural Computing & Applications*, 1–10.

Babakhani, S. M., Bahmani, M., Shariati, J., Badr, K., & Balouchi, Y. (2015). Comparing the capability of artificial neural network (ANN) and CSMHYD program for predicting of hydrate formation pressure in binary mixtures. *Journal of Petroleum Science Engineering*, *136*, 78–87. doi:10.1016/j.petrol.2015.11.002

Baghban, A., Namvarrechi, S., Phung, L. T. K., Lee, M., Bahadori, A., & Kashiwao, T. (2016). Phase equilibrium modelling of natural gas hydrate formation conditions using LSSVM approach. *Petroleum Science and Technology*, *34*(16), 1431–1438. doi:10.1080/10916466.2016.1202966

Blackwell, W. J., & Chen, F. W. (2009). *Neural networks in atmospheric remote sensing*. Artech House.

Bouchard, C., & Granjean, B. P. A. (1995). A neural network correlation for the variation of viscosity of sucrose aqueous solutions with temperature and concentration. *Lebensmittel-Wissenschaft + Technologie*, *28*(1), 157–159. doi:10.1016/S0023-6438(95)80029-8

Cai, L., Pethica, B. A., Debenedetti, P. G., & Sundaresan, S. (2016). Formation of cyclopentane methane binary clathrate hydrate in brine solutions. *Chemical Engineering Science*, *141*, 125–132. doi:10.1016/j.ces.2015.11.001

Cao, J., Zhu, S., Li, C., & Han, B. (2020). Integrating support vector regression with genetic algorithm for hydrate formation condition prediction. *Processes (Basel, Switzerland)*, *8*(5), 1–11. doi:10.3390/pr8050519

Chapoy, A., Mohammadi, A. H., & Richon, D. (2007). Predicting the hydrate stability zones of natural gases using artificial neural networks. *Oil & Gas Science and Technology-Revue de l'IFP*, *62*(5), 701–706. doi:10.2516/ogst:2007048

Chavoshi, S., Safamirzaei, M., & Pajoum Shariati, F. (2018). Evaluation of empirical correlations for predicting gas hydrate formation temperature. *Gas Processing Journal*, *6*(2), 15–36.

Chong, Z. R., Yang, S. H. B., Babu, P., Linga, P., & Li, X.-S. (2016). Review of natural gas hydrates as an energy resource: Prospects and challenges. *Applied Energy*, *162*, 1633–1652. doi:10.1016/j.apenergy.2014.12.061

Collett, T. S., Lewis, R., & Uchida, T. (2000). Growing interest in gas hydrates. *Oilfield Review*, *12*(2), 42–57.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. doi:10.1007/BF00994018

Davie, M. K., Zatsepina, O. Y., & Buffett, B. A. (2004). Methane solubility in marine hydrate environments. *Marine Geology*, *203*(1–2), 177–184. doi:10.1016/S0025-3227(03)00331-1

Elgibaly, A. A., & Elkamel, A. M. (1998). A new correlation for predicting hydrate formation conditions for various gas mixtures and inhibitors. *Fluid Phase Equilibria*, *152*(1), 23–42. doi:10.1016/S0378-3812(98)00368-9

Englezos, P. (1993). Clathrate hydrates. *Industrial & Engineering Chemistry Research*, *32*(7), 1251–1274. doi:10.1021/ie00019a001

Fournaison, L., Delahaye, A., Chatti, I., & Petitet, J.-P. (2004). CO2 hydrates in refrigeration processes. *Industrial & Engineering Chemistry Research*, *43*(20), 6521–6526. doi:10.1021/ie030861r

Ganguly, S. (2003). Prediction of VLE data using radial basis function network. *Computers & Chemical Engineering*, *27*(10), 1445–1454. doi:10.1016/S0098-1354(03)00068-1

Gbaruko, B. C., Igwe, J. C., Gbaruko, P. N., & Nwokeoma, R. C. (2007). Gas hydrates and clathrates: Flow assurance, environmental and economic perspectives and the Nigerian liquified natural gas project. *Journal of Petroleum Science Engineering*, *56*(1–3), 192–198. doi:10.1016/j.petrol.2005.12.011

Ghavipour, M., Ghavipour, M., Chitsazan, M., Najibi, S. H., & Ghidary, S. S. (2013). Experimental study of natural gas hydrates and a novel use of neural network to predict hydrate formation conditions. *Chemical Engineering Research & Design*, *91*(2), 264–273. doi:10.1016/j.cherd.2012.08.010

Giustiniani, M., Tinivella, U., Jakobsson, M., & Rebesco, M. (2013). Arctic ocean gas hydrate stability in a changing climate. *Journal of Geological Research*, (783969), 1–10.

Heydari, A., Shayesteh, K., & Kamalzadeh, L. (2006). Prediction of hydrate formation temperature for natural gas using artificial neural network. *Oil and Gas Business,* 1–10.

Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press. doi:10.7551/mitpress/1090.001.0001

Jain, S. (2017). Introduction to genetic algorithm & their application in data science. *Analytics Vidhya*, 1–31.

Katz, D. L. (1945). Prediction of conditions for hydrate formation in natural gases. *Transactions of the AIME, 160*(1), 140–149. 10.2118/945140-G

Katz, D. L. V. (1959). *Handbook of natural gas engineering*. McGraw-Hill.

Ke, W., & Chen, D. (2019). A short review on natural gas hydrate, kinetic hydrate inhibitors and inhibitor synergists. *Chinese Journal of Chemical Engineering*, *27*(9), 2049–2061. doi:10.1016/j.cjche.2018.10.010

Khamehchi, E., Shamohammadi, E., & Yousefi, S. H. (2013). Predicting the hydrate formation temperature by a new correlation and neural network. *Gas Processing, 1*(1), 41–50. Retrieved from https://www.sid.ir/en/Journal/ViewPaper.aspx?ID=418945

Khan, M. S., Lal, B., Sabil, K. M., & Ahmed, I. (2019). Desalination of seawater through gas hydrate process: An overview. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, *55*(1), 65–73.

Khan, S. H., Kumari, A., Chandrajit, G. D., & Amit, B. M. (2020). Thermodynamic modeling and correlations of CH_4, C_2H_6, CO_2, H_2S, and N_2 hydrates with cage occupancies. *Journal of Petroleum Exploration and Production Technology*, (0123456789). doi:10.1007/s13202-020-00998-y

Koza, J. R. (1994). Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*, *4*(2), 87–112. doi:10.1007/BF00175355

Kumari, A., Hasan, S., Majumder, K. C. B., Arora, A., & Dixit, G. (2021). Physio - chemical and mineralogical analysis of gas hydrate bearing sediments of Andaman Basin. *Marine Geophysical Research*, 1–12. doi:10.1007/s11001-020-09423-9

Kumari, A., Khan, S. H., Misra, A. K., Majumder, C. B., & Arora, A. (2020). Hydrates of Binary Guest Mixtures: Fugacity Model Development and Experimental Validation. *Journal of Non-Equilibrium Thermodynamics*, *45*(1), 39–58. doi:10.1515/jnet-2019-0062

Kumari, A., Madhaw, M., Majumder, C. B., & Arora, A. (2021). Artificial Neural Network: A New Tool for the Prediction of Hydrate Formation Conditions. In Applications of Nature-Inspired Computing in Renewable Energy Systems (pp. 95–115). Academic Press.

Kumari, A., Madhaw, M., Majumder, C. B., Arora, A., & Dixit, G. (2021). *Application of statistical learning theory for thermodynamic modeling of natural gas hydrates*. Petroleum. doi:10.1016/j.petlm.2021.10.005

Kvenvolden, K. A. (1993). *A primer on gas hydrates* (Vol. 1570). US Geological Survey.

Laugier, S., & Richon, D. (2003). Use of artificial neural networks for calculating derived thermodynamic quantities from volumetric property data. *Fluid Phase Equilibria*, *210*(2), 247–255. doi:10.1016/S0378-3812(03)00172-9

Lee, H., Lee, J., Kim, D. Y., Park, J., Seo, Y.-T., Zeng, H., & Moudrakovski, I. L. (2011). *Tuning clathrate hydrates for hydrogen storage. Materials For Sustainable Energy: A Collection of Peer-Reviewed Research and Review Articles from Nature Publishing Group*. World Scientific.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, *11*(2), 431–441. doi:10.1137/0111030

Mehrizadeh, M. (2020). Prediction of gas hydrate formation using empirical equations and data-driven models. *Materials Today: Proceedings*. Retrieved from doi:10.1016/j.matpr.2020.06.058

Mesbah, M., Soroush, E., & Rezakazemi, M. (2017). Development of a least squares support vector machine model for prediction of natural gas hydrate formation temperature. *Chinese Journal of Chemical Engineering, 25*(9), 1238–1248. doi:10.1016/j.cjche.2016.09.007

Mesbah, M., Soroush, E., & Roham, M. (2017). *Phase equilibrium modeling of semi-clathrate hydrates of the CO 2 + H 2 / CH 4 / N 2 + TBAB aqueous solution system*. Academic Press.

Mohammadi, A. H., Belandria, V., & Richon, D. (2010). Use of an artificial neural network algorithm to predict hydrate dissociation conditions for hydrogen+ water and hydrogen+ tetra-n-butyl ammonium bromide+ water systems. *Chemical Engineering Science*, *65*(14), 4302–4305. doi:10.1016/j.ces.2010.04.026

Mohammadi, A. H., Eslamimanesh, A., Richon, D., Gharagheizi, F., Yazdizadeh, M., Javanmardi, J., Hashemi, H., Zarifi, M., & Babaee, S. (2012). Gas hydrate phase equilibrium in porous media: Mathematical modeling and correlation. *Industrial & Engineering Chemistry Research*, *51*(2), 1062–1072. doi:10.1021/ie201904r

Mohammadi, A. H., Martínez-López, J. F., & Richon, D. (2010). Determining phase diagrams of tetrahydrofuran+ methane, carbon dioxide or nitrogen clathrate hydrates using an artificial neural network algorithm. *Chemical Engineering Science*, *65*(22), 6059–6063. doi:10.1016/j.ces.2010.07.013

Mohammadi, A. H., & Richon, D. (2009). Methane hydrate phase equilibrium in the presence of salt (NaCl, KCl, or CaCl2) + ethylene glycol or salt (NaCl, KCl, or CaCl2) + methanol aqueous solution: Experimental determination of dissociation condition. *The Journal of Chemical Thermodynamics*, *41*(12), 1374–1377. doi:10.1016/j.jct.2009.06.012

Mohammadi, A. H., & Richon, D. (2010). Hydrate phase equilibria for hydrogen+ water and hydrogen+ tetrahydrofuran+ water systems: Predictions of dissociation conditions using an artificial neural network algorithm. *Chemical Engineering Science*, *65*(10), 3352–3355. doi:10.1016/j.ces.2010.02.015

Moradi, M. R., Nazari, K., Alavi, S., & Mohaddesi, M. (2013). Prediction of equilibrium conditions for hydrate formation in binary gaseous systems using artificial neural networks. *Energy Technology (Weinheim)*, *1*(2-3), 171–176. doi:10.1002/ente.201200056

Pelckmans, K., Suykens, J. A. K., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., & De Moor, B. (2002). LS-SVMlab: a matlab/c toolbox for least squares support vector machines. Tutorial. KULeuven-ESAT.

Petersen, R., Fredenslund, A., & Rasmussen, P. (1994). Artificial neural networks as a predictive tool for vapor-liquid equilibrium. *Computers & Chemical Engineering*, *18*, S63–S67. doi:10.1016/0098-1354(94)80011-1

Potukuchi, S., & Wexler, A. S. (1997). Predicting vapor pressures using neural networks. *Atmospheric Environment*, *31*(5), 741–753. doi:10.1016/S1352-2310(96)00203-8

Rai, P., Majumdar, G. C., DasGupta, S., & De, S. (2005). Prediction of the viscosity of clarified fruit juice using artificial neural network: A combined effect of concentration and temperature. [Elsevier.]. *Journal of Food Engineering*, *68*(4), 527–533. doi:10.1016/j.jfoodeng.2004.07.003

Rebai, N., Hadjadj, A., Benmounah, A., Berrouk, A. S., & Boualleg, S. M. (2019). Prediction of natural gas hydrates formation using a combination of thermodynamic and neural network modeling. *Journal of Petroleum Science and Engineering, 182*(March), 106270. doi:10.1016/j.petrol.2019.106270

Sablani, S. S., Baik, O.-D., & Marcotte, M. (2002). Neural networks for predicting thermal conductivity of bakery products. *Journal of Food Engineering*, *52*(3), 299–304. doi:10.1016/S0260-8774(01)00119-4

Schmitz, J. E., Zemp, R. J., & Mendes, M. J. (2006). Artificial neural networks for the solution of the phase stability problem. *Fluid Phase Equilibria*, *245*(1), 83–87. doi:10.1016/j.fluid.2006.02.013

Seif, M., & Kamran-pirzaman, A. (2020). Prediction of Gas and Refrigerant Hydrate Equilibrium Conditions With and Without Thermodynamic Inhibitors Using Simple Empirical Correlations. *Journal of Petroleum Science and Technology*, *10*, 61–72.

Sharma, R., Singhal, D., Ghosh, R., & Dwivedi, A. (1999). Potential applications of artificial neural networks to thermodynamics: Vapor–liquid equilibrium predictions. *Computers & Chemical Engineering*, *23*(3), 385–390. doi:10.1016/S0098-1354(98)00281-6

Sinehbaghizadeh, S., Javanmardi, J., Roosta, A., & Mohammadi, A. H. (2019). Estimation of the dissociation conditions and storage capacities of various sH clathrate hydrate systems using effective deterministic frameworks. *Fuel*, *247*, 272–286. doi:10.1016/j.fuel.2019.01.189

Sloan, E. D. (1998). Physical / Chemical Properties of Gas Hydrates. *Gas Hydrates. Relevance to world margin stability and climatic change, 624*(5), 191–196.

Soroush, E., Mesbah, M., Shokrollahi, A., Bahadori, A., & Ghazanfari, M. H. (2014). Prediction of methane uptake on different adsorbents in adsorbed natural gas technology using a rigorous model. *Energy & Fuels*, *28*(10), 6299–6314. doi:10.1021/ef501550p

Soroush, E., Mesbah, M., Shokrollahi, A., Rozyn, J., Lee, M., Kashiwao, T., & Bahadori, A. (2015). Evolving a robust modeling tool for prediction of natural gas hydrate formation conditions. *Journal of Unconventional Oil and Gas Resources*, *12*, 45–55.

Sun, C.-Y., Chen, G.-J., Lin, W., & Guo, T.-M. (2003). Hydrate formation conditions of sour natural gases. *Journal of Chemical & Engineering Data*, *48*(3), 600–602. doi:10.1021/je020155h

Thakur, N. K., & Rajput, S. (2010). *Exploration of gas hydrates: Geophysical techniques*. Springer Science & Business Media.

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*(5), 988–999. doi:10.1109/72.788640 PMID:18252602

Xu, H., Jiao, Z., Zhang, Z., Huffman, M., & Wang, Q. (2021). Prediction of methane hydrate formation conditions in salt water using machine learning algorithms. *Computers & Chemical Engineering*, *151*, 107358. doi:10.1016/j.compchemeng.2021.107358

Yarveicy, H., & Ghiasi, M. M. (2017a). Modeling of gas hydrate phase equilibria: Extremely randomized trees and LSSVM approaches. *Journal of Molecular Liquids*, *243*, 533–541. doi:10.1016/j.molliq.2017.08.053

Yarveicy, H., & Ghiasi, M. M. (2017b). Modeling of gas hydrate phase equilibria: Extremely randomized trees and LSSVM approaches. *Journal of Molecular Liquids, 243*, 533–541. doi:10.1016/j.molliq.2017.08.053

Zahedi, G., Karami, Z., & Yaghoobi, H. (2009). Prediction of hydrate formation temperature by both statistical models and artificial neural network approaches. *Energy Conversion and Management*, *50*(8), 2052–2059. doi:10.1016/j.enconman.2009.04.005

ZareNezhad, B., & Aminian, A. (2012). Accurate prediction of sour gas hydrate equilibrium dissociation conditions by using an adaptive neuro fuzzy inference system. *Energy Conversion and Management*, *57*, 143–147. doi:10.1016/j.enconman.2011.12.021

KEY TERMS AND DEFINITIONS

γ_g: Gas-specific gravity.
AARD%: Average absolute relative deviation.
ANN: Artificial neural network.
ARD%: Absolute relative deviation.
GA: Genetic algorithm.
GP: Genetic programming.
HFT: Hydrate formation temperature.
LSSVM: Least-squares support vector machine.
R^2: Squared correlation coefficient.
RBF: Radial basis function.
RMSD: Root mean square deviation.