energies

MDPI

Article

Development of AI-Based Diagnostic Model for the Prediction of Hydrate in Gas Pipeline

Youngjin Seo [1,†], Byoungjun Kim [2,†], Joonwhoan Lee [3] and Youngsoo Lee [1,*]

[1] Department of Mineral Resource and Energy Engineering, Jeonbuk National University, Jeonju 54896, Korea; yjseo@jbnu.ac.kr
[2] IT Application Research Center, Korea Electronics Technology Institute, Jeonju 54853, Korea; jun0420@keti.re.kr
[3] Division of Computer Science and Engineering, Jeonbuk National University, Jeonju 54896, Korea; chlee@jbnu.ac.kr
* Correspondence: youngsoo.lee@jbnu.ac.kr; Tel.: +82-63-270-2392
† These authors equally contribute to this work.

Abstract: For the stable supply of oil and gas resources, industry is pushing for various attempts and technology development to produce not only existing land fields but also deep-sea, where production is difficult. The development of flow assurance technology is necessary because hydrate is aggregated in the pipeline and prevent stable production. This study established a system that enables hydrate diagnosis in the gas pipeline from a flow assurance perspective. Learning data were generated using an OLGA simulator, and temperature, pressure, and hydrate volume at each time step were generated. Stacked auto-encoder (SAE) was used as the AI model after analyzing training loss. Hyper-parameter matching and structure optimization were carried out using the greedy layer-wise technique. Through time-series forecast, we determined that AI diagnostic model enables depiction of the growth of hydrate volume. In addition, the average R-square for the maximum hydrate volume was 97%, and that for the formation location was calculated as 99%. This study confirmed that machine learning could be applied to the flow assurance area of gas pipelines and it can predict hydrate formation in real time.

Keywords: gas hydrate; diagnostic model; artificial intelligence; stacked auto-encoder; greedy layer-wise

Citation: Seo, Y.; Kim, B.; Lee, J.; Lee, Y. Development of AI-Based Diagnostic Model for the Prediction of Hydrate in Gas Pipeline. *Energies* **2021**, *14*, 2313. https://doi.org/10.3390/en14082313

Academic Editor: Federico Rossi

Received: 23 March 2021
Accepted: 17 April 2021
Published: 20 April 2021

1. Introduction

According to the 2019 EIA Energy Report, global energy consumption will continue to grow until 2050 [1]. According to this report, several resources are used as energy sources, including oil, natural gas, coal, nuclear power, and hydro. Still, the most crucial thing of them is oil and natural gas. Thus, to ensure a stable supply of oil and gas demand, industry is pursuing various technological attempts to effectively develop existing land fields. Hu studied crucial parameters and their effects on recovery factor in the tight reservoir and carbon dioxide adsorption [2], and Mazarei researched stable gas supply in cold weather [3]. This expansion of research leads to the production in areas where production is difficult such as deep seas. In sub-sea production systems, petroleum presenting in the reservoir is a mixture of various components. When pressure and temperature change during production, many deposition problems can occur, such as paraffin, hydrate, and resins. This deposition usually interferes with oil and gas flow in the pipeline, from which the flow assurance concept was proposed [4]. Flow assurance is an engineering technology used to ensure the hydrocarbon fluids are transmitted economically over the life of a project in an inappropriate environment [5].

In general, deposition in inaccessible deep-sea production systems is difficult to eliminate. They usually reduce the cross-sectional area of pipeline and hinder stable

production (Figure 1). For example, there have been 51 cases of wax deposition during production in the Gulf of Mexico over 10 years, and it costs a lot to solve [6].

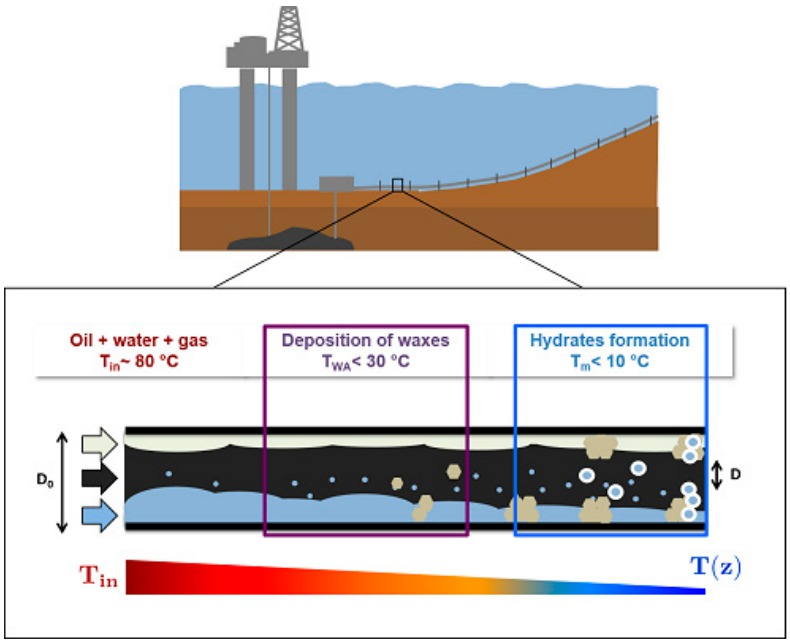

Figure 1. Parametric diagram of the production of paraffin waxes and hydrates in a pipeline (Non-Newtonian Fluid Dynamics Group, 2016).

Moreover, in the production and transportation of natural gas, gas hydrate can be created. Gas hydrate is an ice-like solid compound made from water and gas under low-temperature and high-pressure conditions, known as a nuisance in the gas pipelines [7]. Gas hydrates are frequently occurred in subsea pipelines or permafrost regions. Like paraffin wax, gas hydrate generation causes serious problems because of blocking pipelines (Figure 2). There are several methods to solve the gas hydrate blockage, such as depressurization or inhibitor injection. However, all of these methods are costly and time-consuming. Hydrate formation along the natural gas pipeline has been identified as a serious threat to the success of gas field operation. Annually, a significant operating expense of about hundreds of millions of USD is devoted to hydrate prevention, with half spent on inhibition. In contrast, offshore operations additionally spend approximately USD 1 million per mile on the insulation of subsea pipelines to prevent hydrates [8].

Figure 2. The gas hydrate blockage in pipeline [9].

Even if only one spot of the long pipeline is blocked, the flow rate decreases, and the pipeline is damaged. As a result, the entire gas production system should be shut down for maintenance. Therefore, it is very important to quickly diagnose and gas hydrate formation to prevent pipeline blockage in terms of flow assurance. Accordingly, it is essential to develop a method that can diagnose hydrate deposition as soon as possible. However, it is challenging to predict these depositions for several reasons; the first is that the pipelines are often constructed on the extensive areas of wilderness, grasslands, rivers, sea, and forests, making it difficult to monitor directly. Secondly, kinetic modeling of hydrate formation is very complex. The physics of deposition is very ambiguous and various driving forces may contribute to hydrate formation [10]. Moreover, the thickness and shape of deposition layer are dependent on fluid composition and field environmental conditions. Artificial intelligence (AI)-based methods can overcome these challenges, making it easier to predict hydrate volume.

AI refers to a system in which a computer deduces on its own as if a person is engaging in brain activity, performs professional work through determination, or supports problem solving. The global AI industry is expected to grow exponentially from about USD 600 million in 2016 to USD 36.8 billion in 2025 [11] and to steadily increase in the petroleum industry (Figure 3).

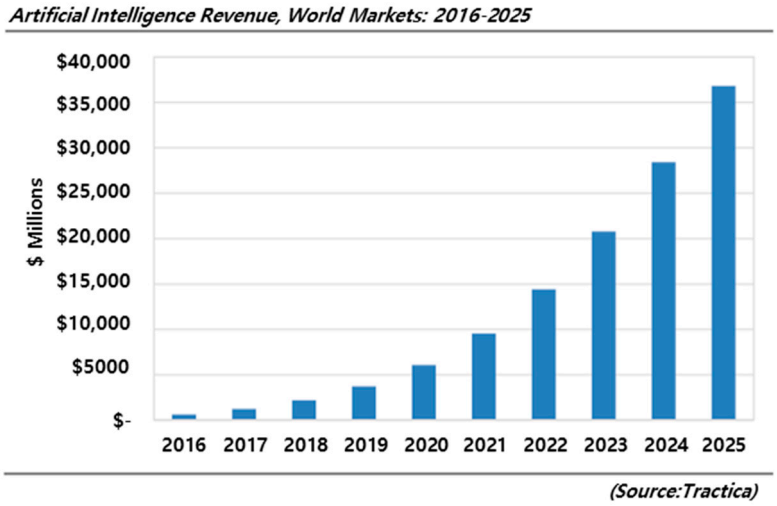

Figure 3. Artificial intelligence revenue, world markets: 2016–2025 [11].

In particular, AI is often used to monitor or predict the deterioration of facilities and is mainly used for the purpose of the flow assurance in oil and gas industry. A mathematical model that could estimate hydrate dissociation conditions based on the feed-forward artificial neural network (ANN) has been presented [12]. Moreover, hydrate formation temperature has been calculated using an ANN [13]. The diagnostic system using AI is highly adaptable to the oil and natural gas field and has great accuracy compared to the existing mathematical models. To utilize the AI model, data that will be used for machine learning is needed. However, because the formation of hydrate usually occurs within the pipeline where monitoring equipment is not installed, there is great difficulty in obtaining actual information from the operation data. The target gas pipeline in this study is installed on the seabed, and it is practically impossible to obtain the hydrate formation data because it is a problem that should never occur during the field operation. The way to overcome this is a digital twin.

A digital twin is a digital replica of a living or non-living physical entity [14]. Digital twin refers to a digital model of potential and actual physical assets, processes, people, places, systems, and devices that can be used for various purposes. Since the digital twin concept was proposed in 2002, it is still emerging as more sophisticated simulations become possible due to IoT, cloud, and big data technologies. The digital twin is actively applied

in heavy industry, which has a relatively high demand, such as producing expensive, complex structure design, long in use, complicated operation to check, and continuous maintenance. As a result, it comprehensively integrated IIoT, AI, big data, and cloud computing technologies; the purpose is to achieve intelligent facility management. In this study, using this digital twin concept, a pipeline flow simulation model was constructed to reflect the target field with operation data, such as pressure, temperature, and the flow rate.

Several studies suggesting gas hydrate diagnostic solutions using AI have common limitations, which are as follows: (1) The models utilized in previous studies are simple models, which require a lot of input but have fewer output values as a result. To predict the hydrate formation in entire pipeline, a large number of output values are needed, such as the pressure, temperature, flow rate, and hydrate volume present in each section. Thus, the prediction of hydrate formation is impossible in all sections of the pipeline. (2) There are no studies that predict hydrate formation within the entire pipeline while simultaneously predicting the growth of hydrates over time. This study attempted to apply AI techniques to establish a pipeline diagnostic solution that can predict the formation of hydrate to mitigate the flow assurance challenges. Notably, the originality of this study is that it can accurately predict the formation and growth of hydrate layer with time at all locations in the pipeline by using machine learning. Moreover, to enhance the field applicability, a model was developed to estimate the pressure and temperature profile, appearance temperature, and deposition thickness throughout the pipeline using the flow rate, pressure, and temperature information, which is generally acquired in the field.

2. Methods

2.1. The Calculation Model of Hydrate Formation Using OLGA

The simulation of multi-phase fluid flow involves conservation equations such as mass, momentum, and energy, and it needs a numerical simulator to solve these equations. The OLGA software has been used in the industry for decades as a predictable multi-phase flow pipeline simulator. OLGA consists of various modules, some of which contain slugging, wax deposition, and hydrate formation. Figure 4 shows the simulation process of OLGA. The input process is needed to define the pipeline materials and network components such as nodes and flow paths. Then the information about the fluid is entered. Finally, boundary and initial conditions are specified. The simulation will be conducted after all the input information has been applied. Three phase model is derived by applying 9 conservation equations. The transport equations are outlined in the following sections in a general, continuous form [15].

In the hydrate case, the hydrate equilibrium curve is used as an input, and the hydrate volume fraction can be predicted. There are 2 options: hydrate check and hydrate kinetic in OLGA. The hydrate check model can be used to obtain information when there is a risk of reaching pressure and temperature conditions under which water can form hydrates. Hydrate kinetic models enable predicting the location where hydrate plugs will be formed in oil and gas pipelines. Especially, a hydrate kinetic model is more suitable for systems with small mass and heat transfer resistance. The hydrate kinetic model includes 3 separate modules, which have hydrate formation, growth, and transportation. Transportation models include mass and thermal diffusion through particle boundary layers and hydrate shells to consider both heat and mass transport limits. The equation of the hydrate formation mechanism is as follows [16].

The driving force of hydrate formation is given by the sub-cooling as follows:

$$\Delta T = T_{ec} - T_{sys} \quad (1)$$

where T_{sys} is the system temperature and T_{ec} is the hydrate equilibrium temperature at the system pressure. The hydrate formation rate is proportional to the temperature driving force. When the driving force is positive and hydrates are forming, the amount of gas consumption due to the hydrate formation is calculated by the reaction rate,

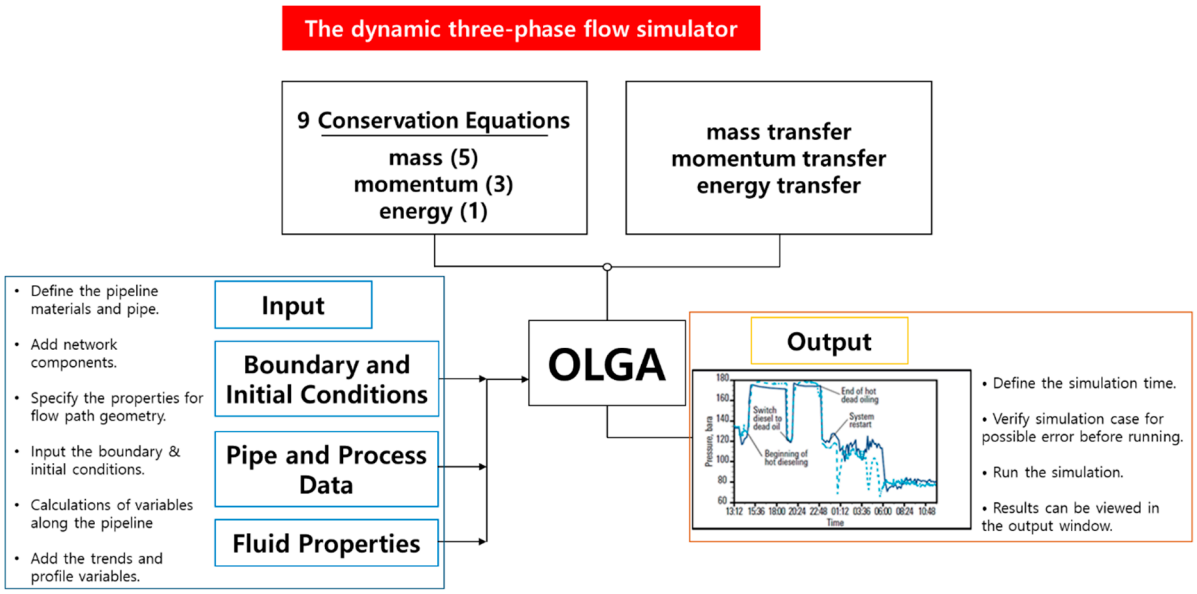

Figure 4. OLGA simulation process (modified by SPT Group, 2017).

$$-r_{gas} = -\frac{dm_{gas}}{dt} = A_s k_1 \exp\left(\frac{k_2}{T}\right)\Delta T \tag{2}$$

where A_s is the surface area between the hydrocarbon-rich phase and the aqueous phase, r_{gas} is the mass of gas consumed per second, and k_1 and k_2 are rate constants.

The agglomerating nature of hydrate particles is addressed using classical crystallization theory, with limited information of hydrate particles, representing the rheological behavior by the relative viscosity. The relative viscosity of suspension with agglomerating particles is given by

$$\mu_r = \frac{1 - \varnothing_{eff}}{\left(1 - \frac{\varnothing_{eff}}{\varnothing_{max}}\right)^2} \tag{3}$$

where $\varnothing_{max}$ is the maximum volume fraction, and effective volume fraction is obtained from

$$\varnothing_{eff} \approx \varnothing\left(\frac{d_A}{d_P}\right)^{(3-f)} \tag{4}$$

The effective volume fraction includes the original volume fraction and the fluid trapped inside the aggregated particle. The symbols of d_P and d_A represent the diameter of monomer particle and aggregated particles, respectively. f is the fractal dimension.

2.2. Machine Learning Using Stacked Auto-Encoder

One of the machine learning techniques, unsupervised learning, is selecting similar cases and learning characteristics using unlabeled data. In particular, this technology can be applied within pipelines that are difficult to obtain data. It can detect weak pipe integrity, such as cracks and corrosion, and analyze the oiliness to prevent risks such as explosions. Moreover, due to the nature of the wide installed area of the pipeline, pressure and temperature drops are hardly detected when the hydrate occurs. Therefore, it is virtually impossible to diagnose them with conventional monitoring methods. In the end, applying artificial intelligence technology is essential, which can reduce the risk of accidents caused by deposits in pipelines and enable rapid response in case of an emergency. In this study, the programming language Python was utilized, and TensorFlow was used as a library for building and operating machine learning programs. After that, the structure of the artificial intelligence model was constructed by applying a stacked auto-encoder. In particular, the fine-tuning technique was used to enable the prediction of hydrate in all

sections of the pipeline by utilizing the pressure, temperature, and flow rates acquired in the oil and gas fields to consider the site applicability.

Auto-encoder is an artificial neural network whose purpose is to fully recover input data from the output layer, similar to a common neural network structure, but characterized by the same size of the input layer and output layer [17]. Figure 5 is the structure of an auto-encoder with a single hidden layer.

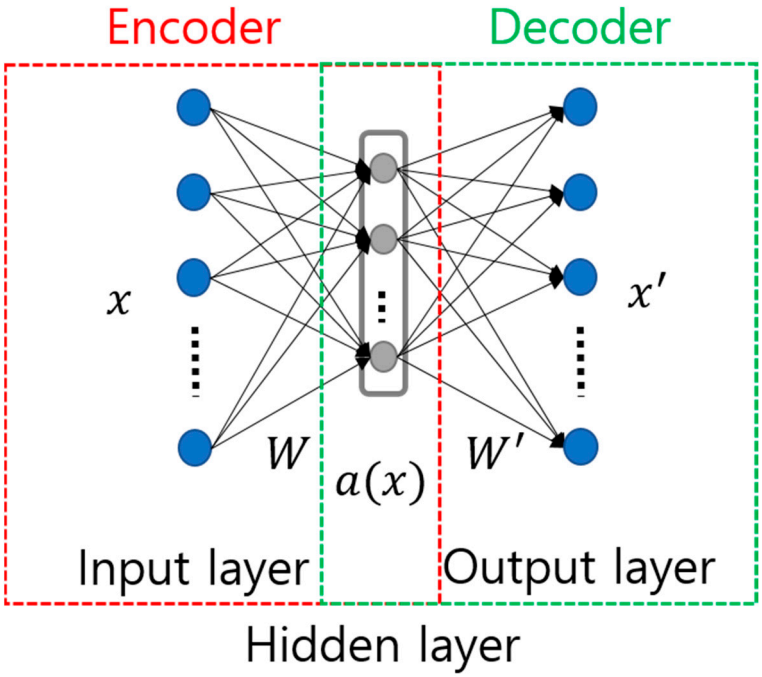

Figure 5. The structure of auto-encoder model with single hidden layer.

As shown in Figure 5, an auto-encoder consists of an encoder and a decoder. The encoder acts to compress the input data into a low-dimensional hidden layer. Simultaneously, the decoder performs the task of restoring the compressed data to the same data level as the input data. Units in the hidden layer need to recover input data from compressed data, implying key features of the input data. Equations (5) and (6) are the formulas calculated from the encoder and decoder of the auto-encoder, respectively.

$$(1)\ \text{Encoding of auto} - \text{encoder}: \ \mathrm{a(x)} = \mathrm{f(Wx + b)} \tag{5}$$

$$(2)\ \text{Decoding of auto} - \text{encoder}: \ x' = \mathrm{f(W'a(x)) + b'} \tag{6}$$

where x and x' are input and reconstructed data, respectively, and the training is carried out to minimize the error of 2 vectors. a is an active function and usually uses nonlinear functions such as Sigmoid, Leak relu, and Tanh.

Stacked auto-encoder (SAE) is a deep-seated neural network model consisting of multiple auto-encoder layers that enable more diverse functions to be expressed in single-layer models [18]. For SAE, greedy layer-wise algorithm is utilized to solve the vanishing gradient problem that occurs when the hidden layer increases [19]. The greedy layer-wise method assumes that there are no other hidden layers in models with multiple hidden layers when learning about the first hidden layer. Later, for the second hidden layer, the first hidden layer parameters are fixed during learning. This method can overcome existing vanishing gradient problems, and complex problems can be solved using multiple hidden layers. In traditional machine learning algorithms, when learning 2 training datasets, 2 models are made independently and use them even if 2 datasets have similar characteristics. On the other hand, transfer learning learns the new model by receiving information from the previously used model. For example, pre-build models contain the

desired pipeline information, such as formation temperature and hydrate volume. Weight and bias calculated by 1:1 matching are stored in a checkpoint. In making the second model, the input variables are changed to obtainable pressure, temperature, and flow rates acquired from the field. The models learned in the first model are brought in to form a new hidden layer in front and conduct secondary learning. It can be used to reduce learning time significantly. At this point, the pre-learned model is defined as pre-training, and the process of using it to learn a new model is defined as fine-tuning.

Most hydrate estimation/prediction studies were conducted using multi-layer perceptron (MLP), but the MLP model is generally used when the input and output dimensions are equal or smaller, and this study had difficulty optimizing and does not guarantee convergence. The time series models LSTM (long short-term memory) and GRU (gated recurrent unit) should consider the time step, making it impossible to predict within all pipeline sections. In this study, the model was selected by comparing and analyzing the training loss values of MLP, LSTM, and SAE (stacked auto-encoder). In general, training loss is an indicator of how stable learning is going, and the lower the better. Figure 6 specifies the results. The MSE (mean square error) values of MLP and LSTM models were increased, which resulted in poor optimization and convergence due to a large number of output data compared to the number of input data.

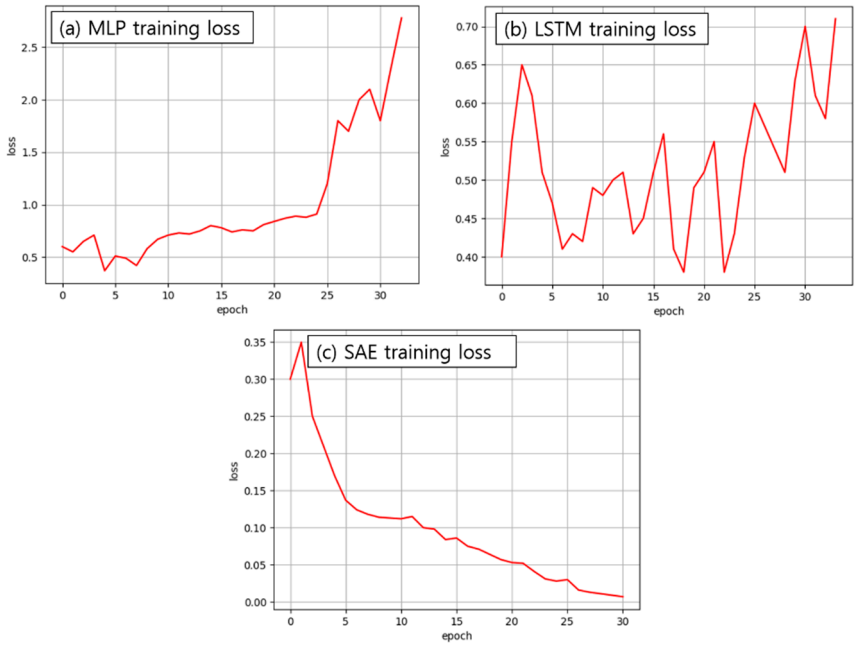

Figure 6. The comparison of training loss according to model differences: (**a**) MLP, (**b**) LSTM, and (**c**) SAE model.

Conversely, SAE is an auto-encoder with multiple hidden layers, configured to mirror encoder and decoder, and used to analyze the characteristics of the data through learning. Additionally, to analyze the characteristics of the output data using the greedy layer-wise technique, we changed the input layer after the pre-learning and relearned by applying the fine-tuning technique. Learning was conducted so that the MSE value was close to zero during learning, which converged on the optimal solution. It was relearned using the analysis of features. Therefore, this study used the SAE model to construct the regression model and construct optimal learning results through repeated experiments.

A library is needed to implement and operate machine learning/deep learning programs, and many libraries such as Torch, Caffe, MXNet, Chainer, and CNTK exist. TensorFlow, which was used in this study, is a machine learning library that provides various functions to easily implement machine learning programs, especially deep learning programs, created by Google. TensorFlow supports various languages such as Python, Java,

and Python is the most advantageous because most useful functions are mostly implemented in Python libraries.

The basic unit of TensorFlow is the calculation graph, and the graph contains the edges representing the tensor and nodes. Each node can provide multiple outputs by multiple inputs [20]. The following Figure 7 is a diagram of the neural network model with multiple hidden layers. The left side is a symbolic diagram of the neural network. The right side is represented by a calculation graph in Tensorboard, using three vectors as input.

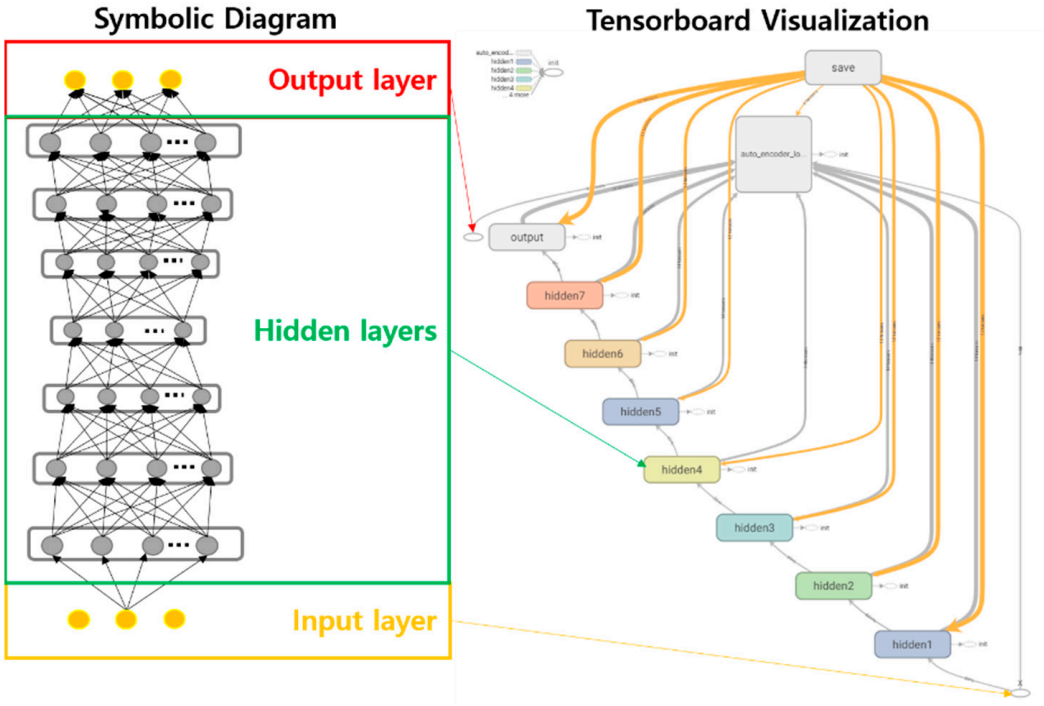

Figure 7. A diagram of a neural network model with multiple hidden layers.

3. Results

3.1. Model Construction and Base Simulation

In this study, a virtual platform was established for the existing offshore gas platform located in Bay of Bengal, offshore western Myanmar, and field data were used to build the model. The schematic of the model is shown in Figure 8.

Each well has a choke valve that controls gas production in the wellhead. The three wellheads are connected to the manifold through a jumper, and the produced gas is transported to the offshore platform through the horizontal subsea pipeline. A topside choke valve is installed in front of the separator. The operating conditions are 60 kg/s of mass flow rate in manifold (20 kg/s for each well), 353.15 K of fluid temperature in the wellhead, and 284.8 K ambient temperature in the inlet of the pipeline. The temperature and pressure conditions in the outlet of the pipeline are 300.45 K and 8.25 MPa, and the thermal conductivity of the pipeline material is 250 W/M^2K.

The pipeline geometry and fluid information are set the same as the field. The horizontal pipeline has 13,000 m length, 0.3174 m of inner diameter, and 5.00×10^{-5} m of roughness. The total pipeline is subdivided into 123 grid sections to improve accuracy. Model specifications are represented in Table 1.

For the target gas field, more than 99% of methane is producing and very little compositions of other components are observed, this model has set up a fluid composed of 100% methane. The hydrate equilibrium curve was derived using CSMHYD, a hydrate phase equilibrium program developed by the Colorado School of Mines (Figure 9).

In general, hydrate occurrence has to be prevented, and continuous inhibitor injection such as KCl and MEG techniques is usually applied to prevent hydrate formation. Therefore, this study assumed that these inhibitor injection lines were failed, and hydrate may

occur. A preliminary study was conducted to make the hydrate forming conditions. The pressure, temperature, and flow rate were set to extreme conditions. As a result, the base simulation case was constructed, and the artificial operation conditions are addressed in Table 2. The results of the base simulation are illustrated in Figure 10a. Hydrate began to occur at around 2000 m, where the fluid temperature dropped below the hydrate formation temperature, and a maximum of 0.75% of the hydrate occurred at 5500 m. The entire simulation period was 1 h, and the results were recorded every 6 min. The growth of hydrate volume fraction is indicated in Figure 10b, and the peak point moved to the backend of the pipeline over time.

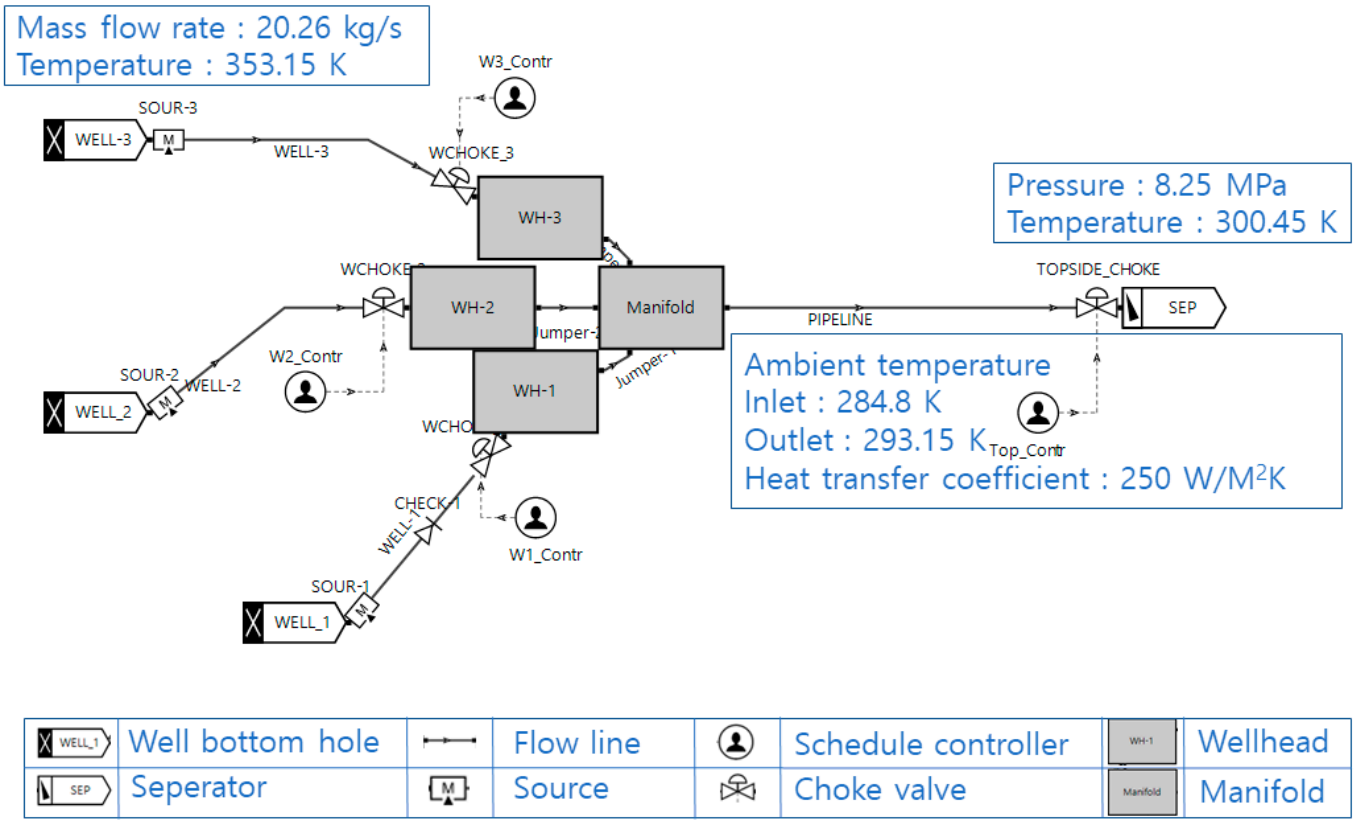

Figure 8. The schematic of gas hydrate model.

Table 1. Model specifications.

Inlet Boundary	
Mass flow rate	60 kg/s
Temperature	353.15 K
Outlet boundary	
Pressure	8.25 MPa
Temperature	300.45 K
Horizontal pipeline	
Ambient temperature	Inlet: 284.8 K Outlet: 293.15 K
Horizontal distance	13,000 m
Heat transfer coefficient	250 W/M^{2}K
Roughness	5.00×10^{-5} m
Inner diameter	0.3174 m
Sections	123

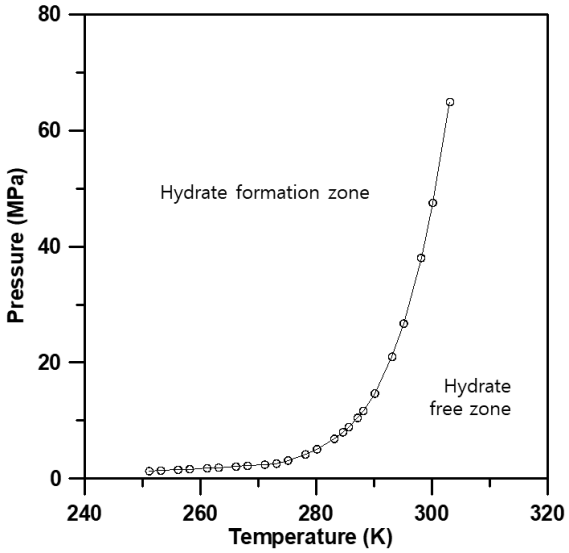

Figure 9. The hydrate equilibrium curve.

Table 2. The difference of field model and base case of hydrate formed.

Operation Constraint	Field Model	Base Case
Mass flow rate (kg/s)	60	60
Inlet fluid temperature (K)	353.15	318.15
Outlet pressure (MPa)	8.25	8.25
Outlet fluid temperature (K)	300.45	300.45
Inlet ambient temperature (K)	284.8	273.15
Heat transfer coefficient (W/M^2K)	250	100

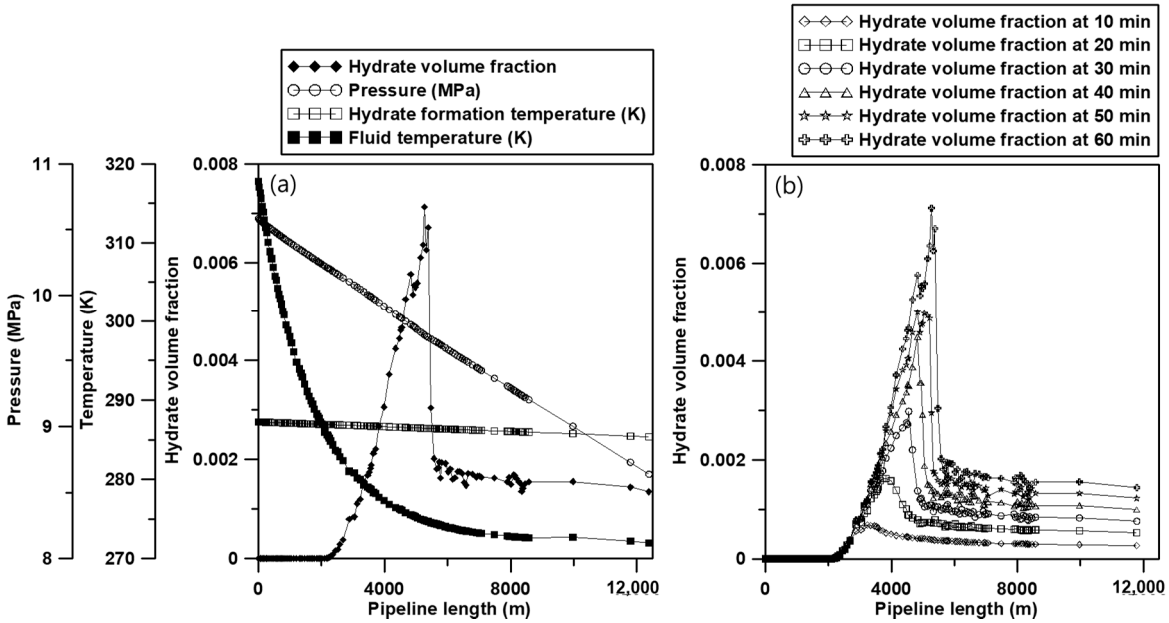

Figure 10. The base model simulation result (**a**) and growth of hydrate volume (**b**).

3.2. Sensitivity Study for the Generation of Learning Data

To make input data for machine learning, different OLGA cases with different locations and volume fractions of hydrate are needed. Moreover, it is necessary to select the influencing parameters that affect hydrate formation and to understand the changes in

locations and volume fractions of the hydrate. Typically, the five factors (presence of free water, low temperature, high operating pressure, flow pattern, and presence of H_2S and CO_2) affect the hydrate formation. For the convenience of analysis, this model assumes that water needed to form hydrates exists in the pipeline. Thus, except for the presence of free water, other factors are examined for understanding the hydrate formation in the pipeline.

After analyzing the flow pattern of all OLGA cases, we found that the flow pattern remained stratified flow and did not change. In addition, the mole fraction of CO_2 and H_2S had little effect on hydrate formation, and these factors were ignored. Therefore, low temperature and high operating pressure were considered.

The modeling parameters linked to the above two factors are mass flow rate, fluid temperature, ambient temperature, and heat transfer coefficient, and Table 3 shows the range of each parameter. In the case of fluid temperature at the inlet, the range was selected between 278.15 and 318.15 K, which was lower than actual field data, in order to create conditions in which more hydrates were formed. For this field model, the length of the pipeline was long enough that it was thought that a temperature gradient in the pipeline would affect the hydrate volume fraction. Because the inlet is located subsea, and the outlet exists at the offshore platform, the difference in ambient temperatures between the inlet and outlet varies due to the seasonal change in platform temperature. The corresponding effect will affect the formation of hydrates. The temperature gradient outside the pipeline is artificially generated by entering different ambient temperatures in the inlet and outlet. The temperature difference between the inlet and outlet was set as 275.15 to 287.15 K. For the heat transfer coefficient, the range of 100 to 200 W/M2K was set to increase the effect of ambient temperature on the fluid temperature inside the pipeline. A total of 400 OLGA cases were run to obtain the learning data.

Table 3. The range of parameters.

Parameters	Range		
	Min	**Mean**	**Max**
Mass flow rate (kg/s)	60 (base)	80	100
Fluid temperature (K)	278.15	298.15	318.15 (base)
Difference of ambient temperature from inlet to outlet (K)	275.15	281.15	287.15
Heat transfer coefficient (W/M^2K)	100 (base)	150	200

A sensitivity study was carried out on how the four selected parameters affect the location and volume fraction of the hydrate using the results of 1 h.

First, an analysis of flow rates was conducted, which is shown in Figure 11a. The forming position of hydrate was not much different, but the larger the flow rate, the more hydrates were generated at the back of the pipeline. In addition, the maximum hydrate volume tended to decrease as the flow rate increased. This was because the high flow rate increased the flow velocity, and the formed hydrate was stripped.

Among the operating condition variables, the most significant affecting factor on the generation of hydrates was temperature. In the case of fluid temperature, the hydrate volume fraction trend was similar, but the lower the temperature, the more hydrates formed at the front of the pipeline (Figure 11b). The gas flowing inside the pipeline initially had a high temperature, and the temperature was decreased during the flow. When the fluid temperature fell below the hydrate formation temperature, hydrates were generated. At the backend of the pipeline, the hydrate volume fraction was almost the same because the fluid temperature decreased to an ambient temperature that all cases showed similar results.

The temperature difference between the inlet and the outlet is shown in Figure 11c. The overall hydrate volume fraction and location were similar, but large temperature differences decreased the amount of hydrate volume generated at the back of the pipeline. This was because the ambient temperature of the outlet was high.

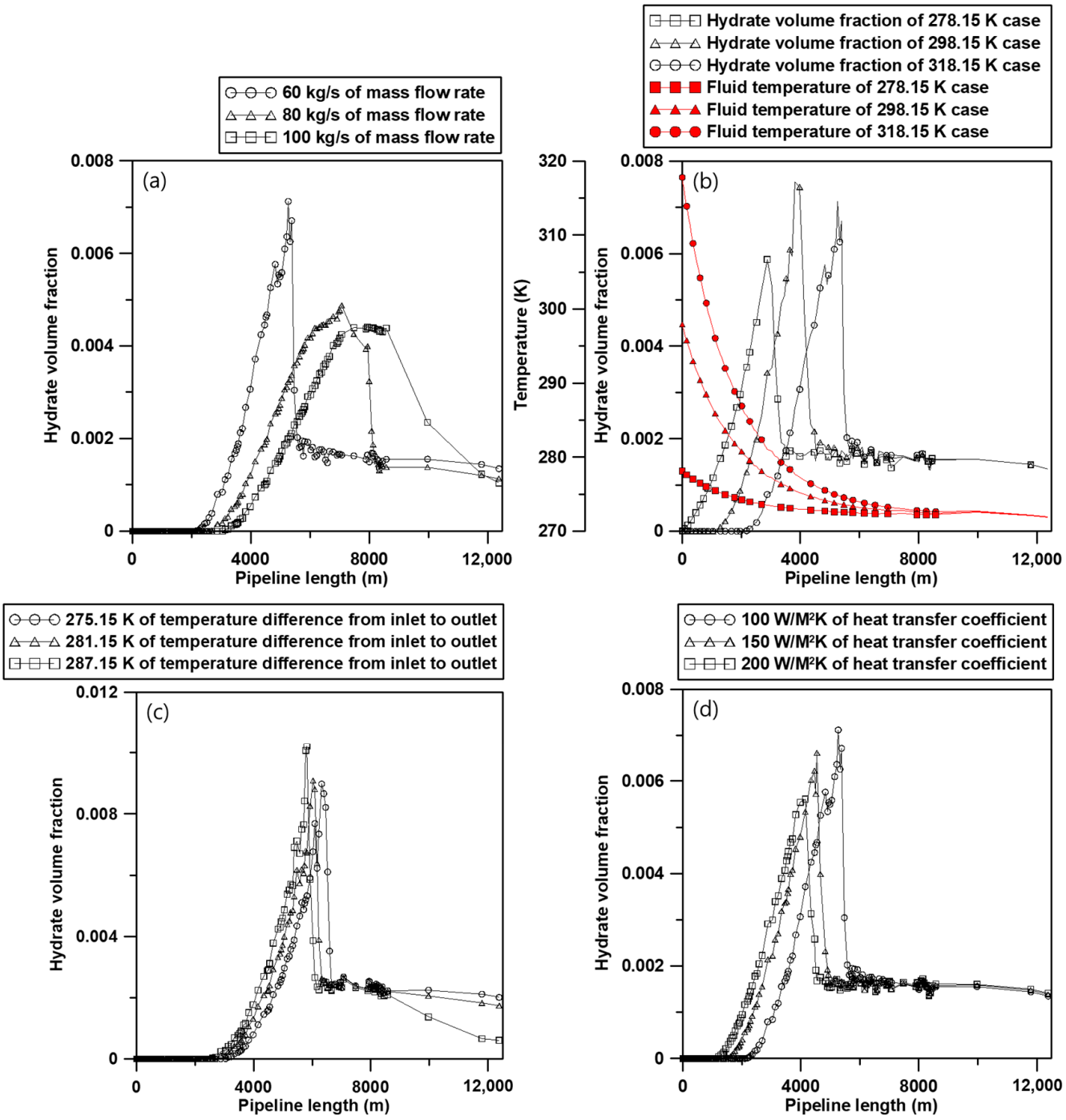

Figure 11. The result of hydrate volume fraction during sensitivity study.

In the heat transfer coefficient, as it increased, the fluid temperature decreased because the effect of ambient temperature became high, and the maximum hydrate volume increased. This tendency was similar to the results of fluid temperature (Figure 11d).

Based on these analyses, we applied a simple random sampling (SRS) technique for obtaining learning data. Each OLGA case consisted of 10 timetables, and each table included the hydrate volume fraction, pressure, temperature, flow rate, and hydrate appearance temperature on 123 grid sections.

3.3. Machine Learning and Validation

Through sensitivity study, location, and volume fraction where hydrate can be formed were analyzed, and learning data were generated for machine learning. The output produced from OLGA was difficult to apply directly to machine learning and required pre-processing. Therefore, the necessary information was extracted from OLGA output, and learning data were processed. TensorFlow was used as a machine learning library, and the SAE was applied as the AI model. The machine learning environment was an I5 core CPU, 64 GB Ram, and GeForce RTX 2080. The model structure is shown in Figure 12.

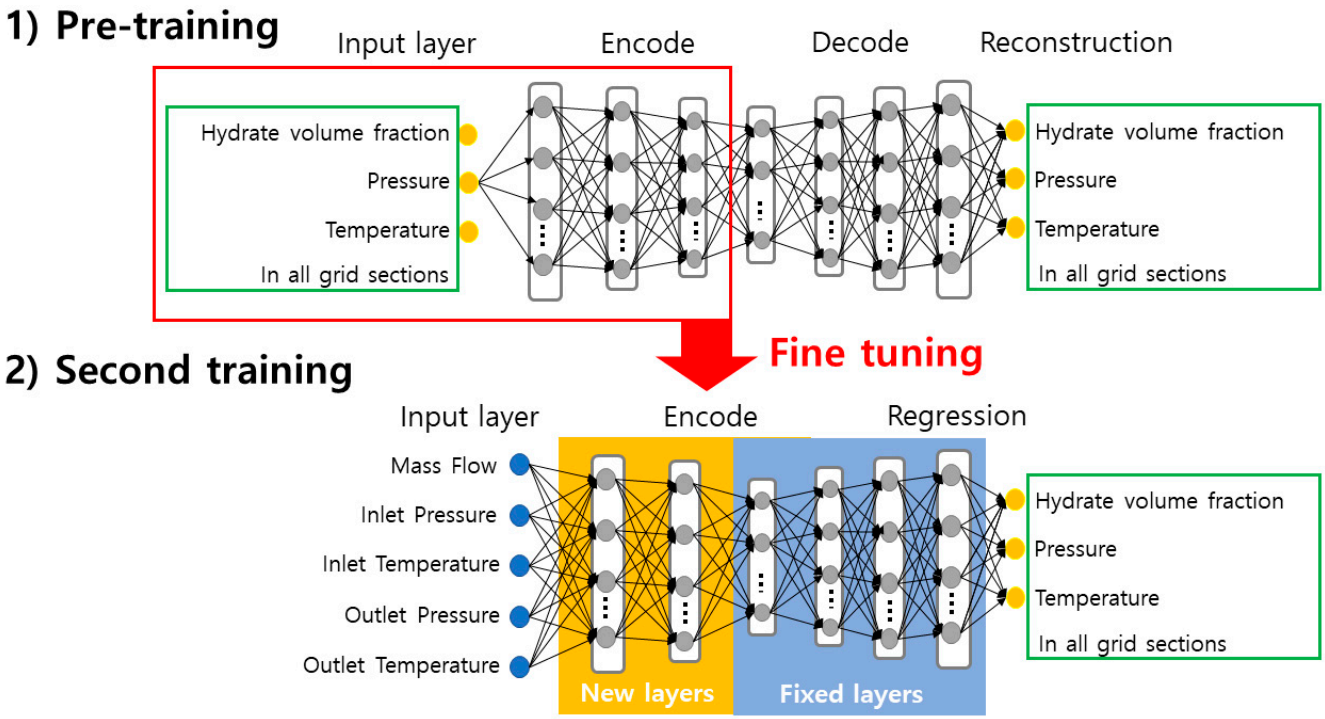

Figure 12. The structure of stacked auto-encoder model for hydrate case.

To establish the model structure, the number of hidden layers and nodes were both hyper-parameters, and therefore we conducted several experiments to find the optimal number. When the number of hidden layers was set as seven at the pre-training phase, the training loss was best converged. In this phase, each layer was trained in an unsupervised manner. As shown in Figure 12, the total 615 data of hydrate volume fraction, pressure, and temperature in all grid sections were set in input and output to proceed pre-training. The reason for learning with the same input and output layers was to analyze and learn about the characteristics of hydrate volume, pressure, and temperature data and to output these data in all grid sections that should ultimately be predicted.

After each layer was learned, fine-tuning was performed to initialize the weight of encoder parts in the pre-trained model. In this process, the weights calculated in the decoder section were frozen because the input dimension changed to five values, but the output value was maintained. Therefore, the weight already obtained in pre-training was reused to reduce learning time in a supervised manner.

The second training phase was also a matching process with new input values, and thus hyper-parameter matching was performed. The number of hidden layers in the encoder was increased gradually, starting from one, and the structure optimization was carried out. As a result, it showed the least convergence value when the number of the hidden layer was one, and when the number of hidden layers was more than two, the difference in convergence values was not significant. Thus, in order to reduce learning time, we utilized a total of six hidden layers, and a total of five inputs (mass flow rate, pressure

and temperature at inlet, pressure and temperature at outlet) were used to recalculate weight through encoders.

The total number of datasets utilized for learning/validation and evaluation was 4000 (10 time tables for each OLGA case), comprising 2,460,000 data (123 sections and 5 variables for each time table). Among them, 3200 datasets (2560 for learning, 640 for validation) were used for learning, while the remaining 800 datasets were used for evaluation. Thereafter, learning data was normalized to the [−1,1] range. The reason for normalization is that the size of the data depends on the variables when the original values are used, which can lead to optimization problems or lack of convergence. The activation function for the hidden layer was configured as Leak relu, and the Tanh was applied as a last output activation function. The basic reason for using the activation function was to maximize the learning effect in AI models with many hidden layers by changing the data to nonlinear. The reason for transforming data into nonlinear was to overcome the limitations of linear classifiers. If the data are linear, no matter how deep the hidden layer is, the linear characteristic such as f(ax + by) = af(x) + bf(y) eventually results in the same outcome when using one hidden layer or dozens of hidden layers. Therefore, a nonlinear activation function was used to take advantage of the use of many hidden layers. In the case of Leak relu, the learning speed is the fastest compared to other activation functions. Moreover, the use of sequential data is likely to result in a vanishing gradient problem. Vanishing gradient is a phenomenon in which the gradient is zero in learning through the hidden layers and no longer learning. Tanh is often used to prevent this.

The learning algorithm (optimization algorithm/Optimizer) was used for optimization. Optimization means finding a factor in the model that minimizes the loss function. The most representative technique is the stochastic gradient descent method (SGD), and the recent development of machine learning technology has led to the development of many optimizers. The most popular optimizer currently used is Adam, a technique combining momentum and the RMSProp method, and an optimizer that improves accuracy and learning step size. Therefore, Adam Optimizer was applied in this study.

Overfitting protection was prepared using Dropout techniques and L2 regularization (7), but if the data order within the epoch is the same, the gradient value in a single batch may not represent the entire dataset, and local minimum or overfitting is likely to occur. To solve this problem, we applied a shuffle-batch to the learning data on every epoch to prevent the gradient falling into zero at local minimum point. The MSE (mean squared error), a qualitative measure of statistical estimated accuracy, was used as a cost function to determine the difference between the estimated value and the actual value. The parameters used for learning are given in Table 4. The batch size and learning rate were calculated through repeated experiments. The epoch was fixed at 10,000, and then the learning was carried out. If the training loss was converged, the learning was finished. Moreover, grid search was performed several times for weight optimization.

$$\mathrm{L(x,y)} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - h_\theta(x_i)\right)^2 + \lambda\sum_{i=1}^{n}\theta_i^2 \tag{7}$$

where n is the number of elements, y_i and $h_\theta(x_i)$ are real value and predicted value, λ is learning rate, and θ_i is weight.

In the pre-training phase, in order to output the values that should ultimately be predicted, we set the total 369 data (123 sections × 3 variables) of hydrate volume fraction, pressure, and temperature in the input–output layer. A total of 80 iterative grid searches in pre-training determined the optimal number of nodes in hidden layers with the least value of MSE. Table 5 represents an example of grid search. The index means the count of grid search. Layer means the hidden layer, and the numbers in the table mean the number of nodes of each hidden layer. As a result of the grid search, the MSE value of index 79 showed the lowest value of 0.98×10^{-3}. Figure 13 shows the training and validation process of index 79, and the training loss converged at epoch 1000. Therefore, the decoder

structure, the number of nodes, and the weights in hidden layers 4 to 7 were frozen to shorten the learning time of second training.

Table 4. The parameters used for machine learning.

Input	615 (pre-training)/5 (second training)
Output	615
Learning rate	1.00×10^{-6}
Batch size	1560
Epoch	10,000
Drop out	0.6
L2 regularization	0.50×10^{-3}
Activation function	Leak relu/Tanh
Optimizer	Adam

Table 5. The result of grid search about pre-training.

Index	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 7	MSE
1	128	64	32	16	32	64	128	32.78
2	256	128	64	32	64	128	256	14.78
3	1024	512	256	128	256	512	1024	1.59
4	2048	1024	512	256	512	1024	2048	2.96
5	512	256	128	64	128	256	512	3.08
...	...	...	...	...	...	...	...	...
79	2048	1024	512	256	512	1024	2048	0.98×10^{-3}
80	4096	2048	1024	512	1024	2048	4096	1.00

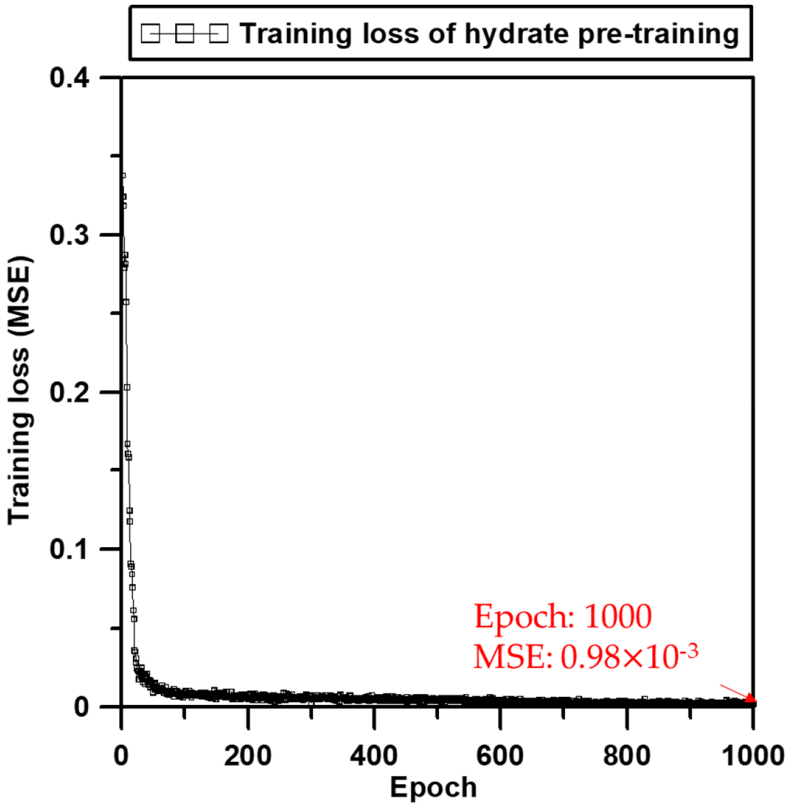

Figure 13. Training loss of index 79 in pre-training of hydrate.

In the second training phase, new hidden layers were added to the encoder part to use the input values mass flow rate, pressure and temperature in inlet, and pressure and temperature in outlet. For the optimization of the AI model, we conducted 80 grid searches again. Table 6 shows that the structure of index 6 had the smallest MSE. Therefore, this model was selected as the final optimized model. Moreover, Figure 14 shows the training and validation process of index 6, wherein the training loss converged with an MSE value of 1.04×10^{-3} at epoch 300.

Table 6. The result of grid search about second training.

Index	Layer 1	Layer 2	Layer 4 (Fixed)	Layer 5 (Fixed)	Layer 6 (Fixed)	Layer 7 (Fixed)	MSE
1	16	32	512	256	1024	2048	1.85×10^{-3}
2	16	64					1.74×10^{-3}
3	32	128					1.09×10^{-3}
4	64	256					1.12×10^{-3}
5	128	64					1.28×10^{-3}
6	128	256					1.04×10^{-3}
...	...	...					...
79	256	256					1.08×10^{-3}
80	256	512					1.07×10^{-3}

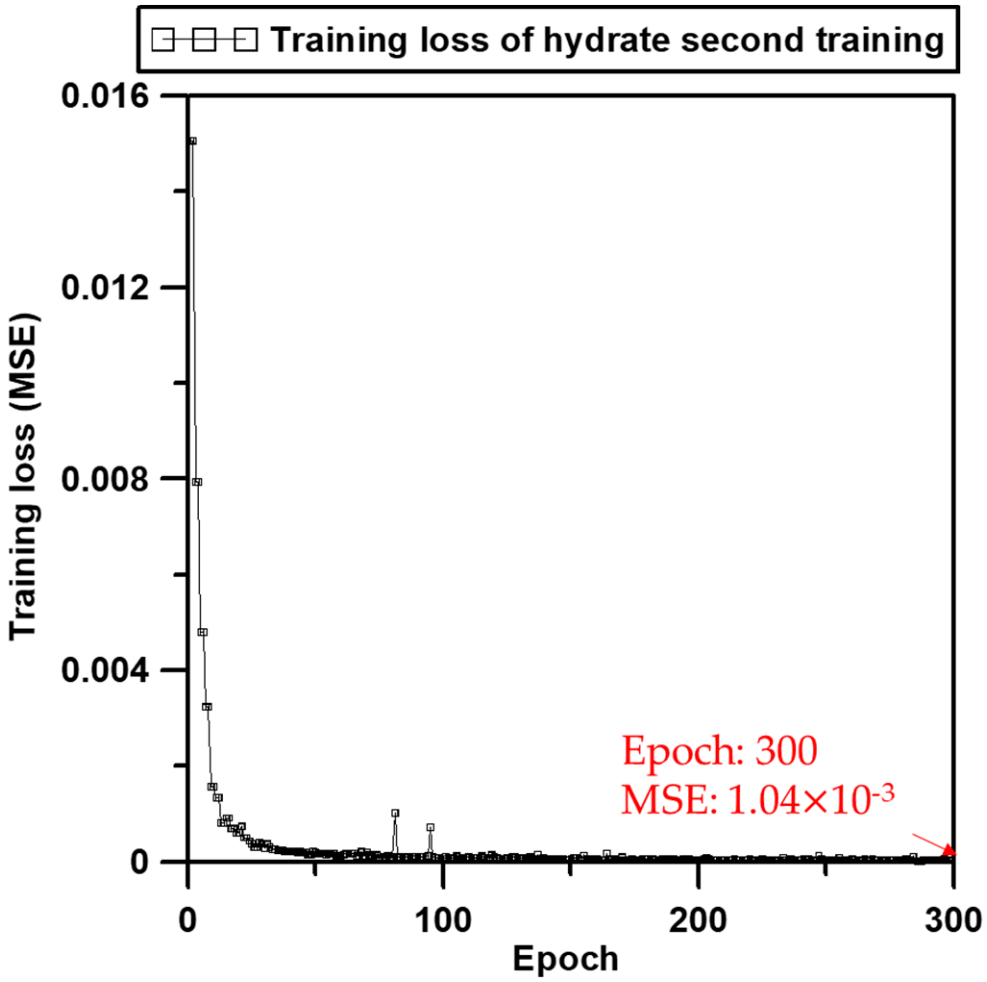

Figure 14. Training loss of index 6 in a second training of hydrate.

4. Discussion

The optimal model was constructed through learning and validation, and the model was evaluated using the test dataset with the R-square method. It was calculated through a comparison between actual data and prediction. This value ranges from zero to one, and the model is useful if the coefficient of determination has a value close to one.

A total of 800 datasets (80 OLGA cases) were used for evaluation. The evaluation dataset had mass flow rate and pressure, temperature, and hydrate volume fraction at each grid section. The mass flow rate, pressure, and temperature in inlet, pressure, and temperature in the outlet of evaluation data were received as input signals. The result value derived from the AI model, the hydrate volume fraction, pressure, and temperature were compared with the value of evaluation data.

To ensure that the growth of the hydrate was properly predicted when the input signal was continuously entered, the hydrate diagnostic model carried out a time-series forecast, which evaluated accuracy over time in a single OLGA case, and an accuracy evaluation for the location where the maximum volume was formed. Table 7 and Figure 15 show the result of the time-series forecast for one OLGA case. In general, pressure and fluid temperature showed relatively high accuracy. In Figure 15, it is possible to describe the growth of hydrate volume, and it was shown to have an accuracy of 77% of hydrate volume. In terms of hydrate volume, the accuracy was low in the early times when the hydrate volume was small, but the accuracy increased as the volume of hydrate grew. The actual and predicted data in Figure 15 were rearranged on the basis of the maximum hydrate volume to prove the logic that accuracy increased as the hydrate volume increased (Figure 16). According to the figures, the predicted accuracy also increased as the maximum hydrate volume increased over time. Therefore, as the hydrate volume increased over time, the prediction accuracy increased.

Table 7. The result of R-square evaluation for one OLGA case.

R-Square Evaluation Result			
Time-Series	**Maximum Hydrate Volume in all Grid Sections**	**Pressure in all Grid Sections**	**Fluid Temperature in all Grid Sections**
6 min	0.46	0.99	0.94
12 min	0.59	1	0.92
18 min	0.68	0.99	0.95
24 min	0.78	1	0.92
30 min	0.80	0.99	0.89
36 min	0.84	1	0.94
42 min	0.86	0.99	0.94
48 min	0.87	0.98	0.89
54 min	0.89	1	0.91
60 min	0.95	1	0.87
Average	0.77	0.99	0.91

Through the previous time-series forecast results, we were able to determine that prediction accuracy for the maximum hydrate volume and location was high at a later time. However, it was confirmed that the maximum hydrate volume at 60 min was slightly different from the formed location. Therefore, an overall performance analysis was performed for the location and maximum hydrate volume. A total of 80 evaluation datasets were extracted and compared with AI diagnosis results. Table 8 shows the maximum hydrate volume and formation position in each case of the evaluation dataset and shows the prediction accuracy. According to Figure 17, the R-square value of the maximum hydrate volume was measured as 97%, and it was found that the volume of hydrate was overestimated below 0.6% and underestimated above. For the formed location, the value of R-square was 99%, but the error was calculated through the MAE because the error range of the actual data and forecasts existed. On average, there was a location error of about 261 m. For this pipeline model, 13,000 m pipeline was divided into 123 grid sections, and

thus each section was approximately 105 m long. Considering the number of grid sections, we found that the error was not significant.

The diagnostic system was constructed for detecting hydrate formation (Figure 18a). When mass flow rate, pressure, and temperature in inlet, pressure, and temperature in the outlet are delivered through the input module, it is transferred to the input data of the AI model in real-time. In the AI model, the hydrate volume, pressure, and temperature in the pipeline are predicted over time by utilizing input data. The predicted results are illustrated in the GUI program over time, which is shown in Figure 18b. When hydrate is formed, the warning message is printed, and, finally, all records are stored in the CSV form.

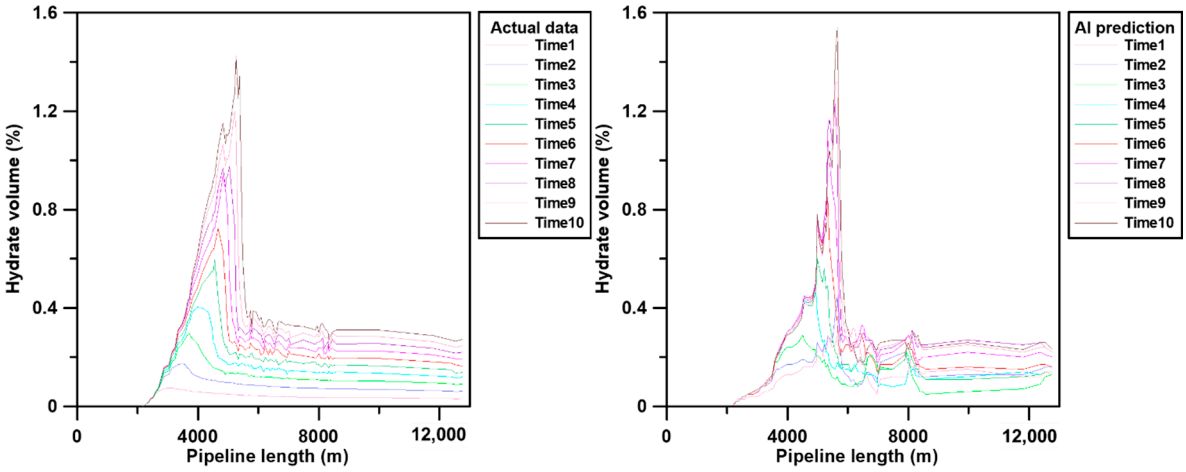

Figure 15. The difference of actual data and AI prediction results.

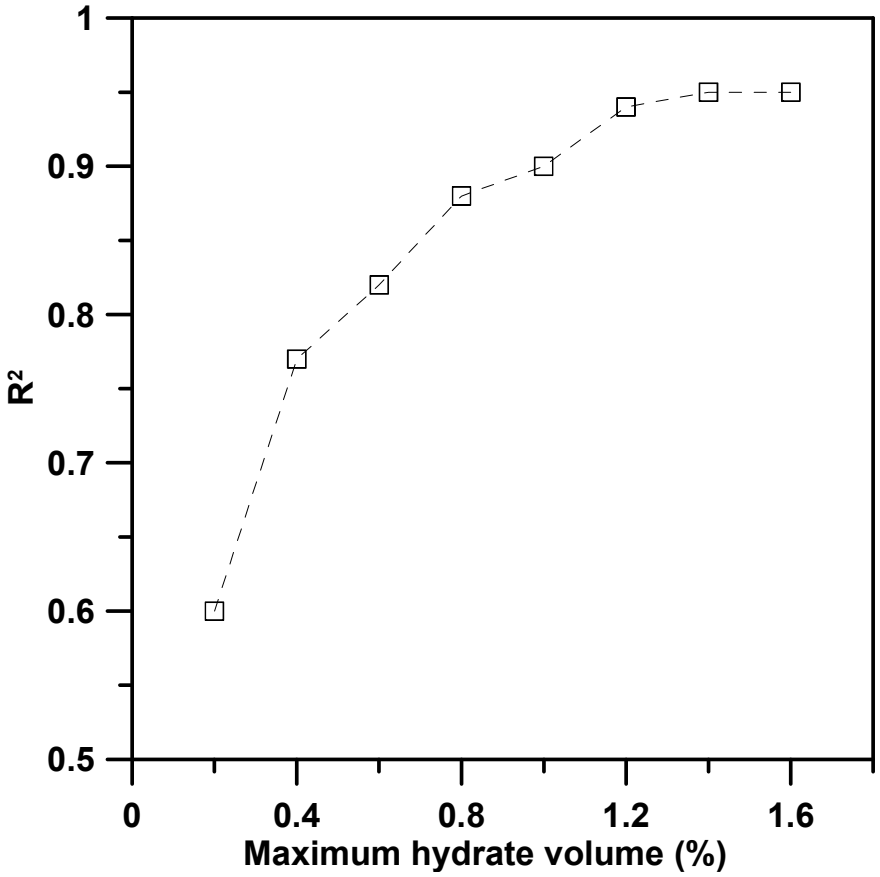

Figure 16. The results of R-square according to maximum hydrate volume.

Table 8. The accuracy of maximum hydrate volume and formed location at 60 min about total 80 OLGA cases using R-square and MAE.

OLGA Cases	Actual Maximum Hydrate Volume (%)	Predicted Maximum Hydrate Volume (%)	Actual Location of Maximum Hydrate Volume (m)	Predicted Location of Maximum Hydrate Volume (m)
Case1	0.41	0.48	8048	8082
Case2	0.69	0.65	10,451	11,795
Case3	0.31	0.49	9432	9970
Case4	0.47	0.48	6781	6781
Case5	0.43	0.49	11,991	11,795
Case6	0.70	0.68	916	940
Case7	0.41	0.41	12,171	12,381
Case8	0.73	0.68	9456	9970
Case9	1.06	1.00	11,939	11,795
Case10	1.20	1.20	6958	6958
		...		
Case71	1.00	1.00	8237	8259
Case72	0.48	0.49	12,201	11,795
Case73	1.18	1.18	4985	4985
Case74	0.98	1.00	12,413	12,381
Case75	0.65	0.60	2296	2291
Case76	0.73	0.68	10,435	9970
Case77	0.70	0.63	8577	8082
Case78	1.16	1.18	10,835	9970
Case79	0.21	0.49	2823	2877
Case80	0.62	0.62	8632	8082
Average	R-square = 0.97		MAE = 261	

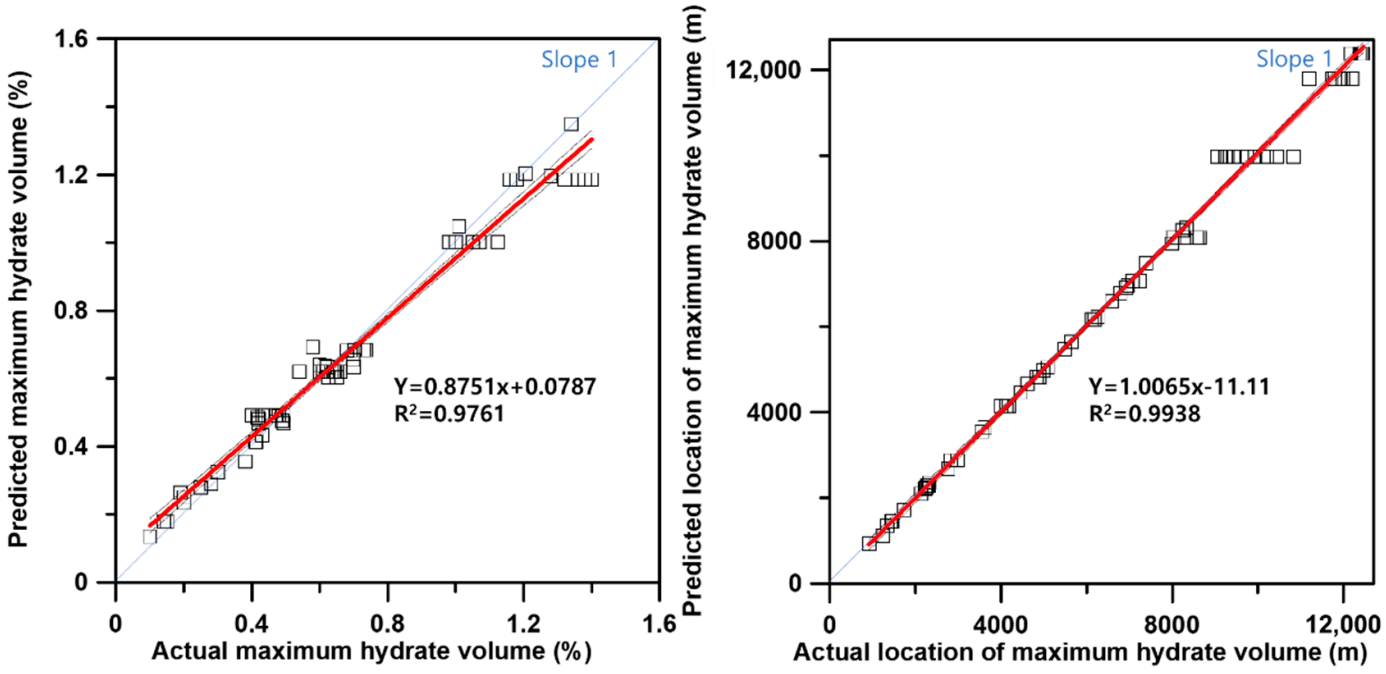

Figure 17. The accuracy of maximum hydrate volume and location with 80 evaluation cases.

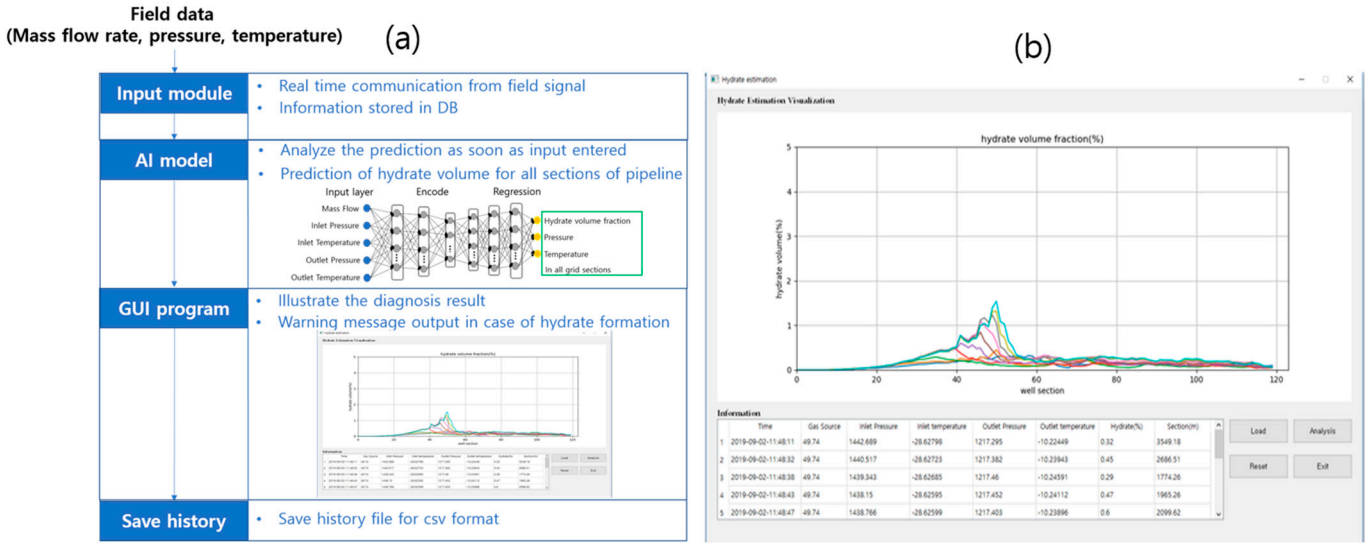

Figure 18. The flow chart (**a**) and GUI program (**b**) of diagnostic system for detecting hydrate formation.

5. Conclusions

In this study, a machine learning technique was applied to establish a system that can diagnose hydrate for flow assurance purposes in gas pipelines. A parametric and sensitivity study was conducted to identify the formation characteristics of hydrate. Based on these results, we generated and evaluated learning data for machine learning. Hyperparameter matching and structure optimization were carried out using the SAE model and the greedy layer-wise technique. The detailed procedures and results are as follows.

(1) Through time-series forecast, we determined that AI diagnostic models could depict growth of hydrate volume. As the hydrate volume increased, the AI model was able to diagnose more accurately. Moreover, the predicted trend for the entire pipeline was more similar to the actual data.

(2) In the evaluation of overall performance, the average R-square for the maximum hydrate volume was 97%, and that for the formation position was calculated as 99%.

(3) The developed AI model can indicate abnormalities within a very short time and accurately diagnose maximum hydrate volume and formation location. This study confirmed that AI could be applied to the flow assurance area of petroleum pipelines. It is expected that this study can be applied to the pipelines lying in various environments.

Author Contributions: Conceptualization, Y.L. and J.L.; software, Y.S.; validation, Y.S. and B.K.; writing—original draft preparation, Y.S. and B.K.; writing—review and editing, Y.S. and Y.L.; supervision, J.L. and Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Energy Efficiency & Resources Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (no. 20172510102150).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2020-0-01557).

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

AI	artificial intelligence
ANN	artificial neural network
ΔT	sub-cooling
T_{sys}	system temperature
T_{ec}	hydrate equilibrium temperature at system pressure
r_{gas}	mass of gas consumed per second
A_s	surface area
k_1	rate constant
k_2	rate constant
μ_r	relative viscosity
$\varnothing_{max}$	maximum volume fraction
$\varnothing_{eff}$	effective volume fraction
d_P	diameter of monomer particle
d_A	diameter of aggregated particle
f	fractal dimension
x	input
x'	reconstructed data
W	weight
b	bias
SAE	stacked auto-encoder
MLP	multi-layer perceptron
LSTM	long short-term memory
GRU	gated recurrent unit
MSE	mean square error
SRS	simple random sampling
SGD	stochastic gradient descent
n	number of elements
y_i	real value
$h_\theta(x_i)$	predicted value
λ	learning rate
θ_i	weight in L2 regularization

References

1. Osaki, K. US Energy Information Administration (EIA): 2019 Edition US Annual Energy Outlook report (AEO2019). *Haikan Gijutsu* **2019**, *61*, 32–43.
2. Hu, X.; Xie, J.; Cai, W.; Wang, R.; Davarpanah, A. Thermodynamic effects of cycling carbon dioxide injectivity in shale reservoirs. *J. Pet. Sci. Eng.* **2020**, *195*, 107717. [CrossRef]
3. Mazarei, M.; Davarpanah, A.; Ebadati, A.; Mirshekari, B. The feasibility analysis of underground gas storage during an integration of improved condensate recovery processes. *J. Pet. Explor. Prod. Technol.* **2018**, *9*, 397–408. [CrossRef]
4. Brower, D.; Prescott, C.; Zhang, J.; Howerter, C.; Rafferty, D. Real-Time Flow Assurance Monitoring with Non-Intrusive Fiber Optic Technology. In Proceedings of the Offshore Technology Conference, Houston, TX, USA, 2–5 May 2005.
5. Bai, Y.; Bai, Q. *Subsea Engineering Handbook*; Gulf Professional Publishing: Houston, TX, USA, 2018.
6. Wood, D.; Mokhatab, S. Gas monetization technologies remain tantalizingly on the brink. *World Oil* **2008**, *229*, 103–108.
7. Menon, E.S. *Gas Pipeline Hydraulics*; CRC Press: Boca Raton, FL, USA, 2005.
8. Jassim, E.; Abdi, M.A.; Muzychka, Y. A new approach to investigate hydrate deposition in gas-dominated flowlines. *J. Nat. Gas Sci. Eng.* **2010**, *2*, 163–177. [CrossRef]
9. Makwashi, N.; Zhao, D.; Ismaila, T.; Paiko, I. Pipeline Gas Hydrate Formation and Treatment: A Review. In Proceedings of the 3rd National Engineering Conference on Building the Gap between Academia and Industry, Faculty of Engineering, Bayero University, Kano, Nigeria, 31 May 2018.
10. Foroozesh, J.; Khosravani, A.; Mohsenzadeh, A.; Mesbahi, A.H. Application of artificial intelligence (AI) in kinetic modeling of methane gas hydrate formation. *J. Taiwan Inst. Chem. Eng.* **2014**, *45*, 2258–2264. [CrossRef]
11. Tractica. *Artificial Intelligence Market Forecasts*; Tractica: Boulder, CO, USA, 2016.
12. Mohammadi, A.H.; Belandria, V.; Richon, D. Use of an artificial neural network algorithm to predict hydrate dissociation conditions for hydrogen+water and hydrogen+tetra-n-butyl ammonium bromide+water systems. *Chem. Eng. Sci.* **2010**, *65*, 4302–4305. [CrossRef]

13. Zahedi, G.; Karami, Z.; Yaghoobi, H. Prediction of hydrate formation temperature by both statistical models and artificial neural network approaches. *Energy Convers. Manag.* **2009**, *50*, 2052–2059. [CrossRef]
14. El Saddik, A. Digital Twins: The Convergence of Multimedia Technologies. *IEEE MultiMedia* **2018**, *25*, 87–92. [CrossRef]
15. SPT Group. *OLGA 2017, User Manual, Dynamic Multiphase Flow Simulator*; SPT Group: Houston, TX, USA, 2017.
16. Turner, D.; Boxall, J.; Yang, S.; Kleehammer, D.; Koh, C.; Miller, K.; Sloan, E.; Xu, Z.; Matthews, P.; Talley, L. Development of a hydrate kinetic model and its incorporation into the OLGA2000®transient multiphase flow simulator. In Proceedings of the 5th international conference on gas hydrates, Trondheim, Norway, 12–16 June 2005; pp. 12–16.
17. Hinton, G.E.; Zemel, R.S. Autoencoders, minimum description length, and Helmholtz free energy. *Adv. Neural Inf. Process. Syst.* **1994**, *6*, 3–10.
18. Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.-A. Extracting and composing robust features with denoising auto-encoders. In Proceedings of the 25th international conference on Machine learning, Helsinki, Finland, 5–9 July 2008; pp. 1096–1103.
19. Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. *Adv. Neural Inf. Process. Syst.* **2007**, *19*, 153.
20. Vishnu, A.; Siegel, C.; Daily, J. Distributed tensorflow with MPI. *arXiv* **2016**, arXiv:1603.02339.