applied sciences

MDPI

Article

Generation of Synthetic Compressional Wave Velocity Based on Deep Learning: A Case Study of Ulleung Basin Gas Hydrate in the Republic of Korea

Minsoo Ji [1], Seoyoon Kwon [1], Min Kim [2], Sungil Kim [3] and Baehyun Min [1,2,*]

[1] Department of Climate and Energy Systems Engineering, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea
[2] Severe Storm Research Center, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea
[3] Marine Geology & Energy Division, Petroleum Energy Research Center, Korea Institute of Geoscience and Mineral Resources, 124, Gwahak-ro, Yuseong-gu, Daejeon 34132, Korea
* Correspondence: bhmin01@ewha.ac.kr; Tel.: +82-2-3277-6946

Abstract: This study proposes a deep-learning-based model to generate synthetic compressional wave velocity (Vp) from well-logging data with application to the Ulleung Basin Gas Hydrate (UBGH) in the East Sea, Republic of Korea. Because a bottom-simulating reflector (BSR) is a key indicator to define the presence of gas hydrate, this study generates the Vp for identifying the BSR by detecting the morphology of the hydrate in terms of the change in acoustic velocity. Conventional easy-to-acquire logging parameters, such as gamma-ray, neutron porosity, bulk density, and photoelectric absorption, were selected as model inputs based on a sensitivity analysis. Long short-term memory (LSTM) and an artificial neural network (ANN) were used to design an efficient learning-based predictive model with sensitivity analysis for hyperparameters. The LSTM model outperforms the ANN model by preserving the geological sequence of the well-logging data. Ten-fold cross-validation was conducted to verify the consistency of the LSTM model and yielded satisfactory results, with an average coefficient of determination greater than 0.8. These numerical results imply that generating synthetic well-logging via deep learning can accurately estimate missing well-logging data, contributing to the reservoir characterization of gas-hydrate-bearing sediments.

Keywords: deep learning; compressional wave velocity; well-logging; Ulleung Basin Gas Hydrate

Citation: Ji, M.; Kwon, S.; Kim, M.; Kim, S.; Min, B. Generation of Synthetic Compressional Wave Velocity Based on Deep Learning: A Case Study of Ulleung Basin Gas Hydrate in the Republic of Korea. *Appl. Sci.* **2022**, *12*, 8775. https://doi.org/10.3390/app12178775

Academic Editors: Xianzhi Song, Hangyu Li and Shuyang Liu

Received: 26 July 2022
Accepted: 29 August 2022
Published: 31 August 2022

1. Introduction

Methane gas hydrate is an ice-like crystalline solid composed of methane gas and water molecules formed under low temperature and high pressure [1]. Because of its high gas-storage capacity, gas hydrate has been considered a clean and abundant new energy resource [2–6]. Because of the temperature and pressure conditions required for its presence, gas hydrate has been discovered primarily in permafrost and continental boundaries worldwide. Its global reserve is estimated in the range of 1×10^{15} to 120×10^{15} m^3 [7–9]. Because hydrate-bearing sediments are less consolidated, hydrate formation properties such as porosity differ significantly from those of conventional sandstone and carbonate reservoirs. A bottom-simulating reflector (BSR) is a key indicator to define the presence of gas hydrate. In addition to a seismic log, it is essential to identify the BSR to obtain an acoustic log, measured by compressional wave velocity (Vp), and the shear wave velocity (Vs), by detecting the morphology of the hydrate in terms of the change in acoustic velocity [10]. However, it is challenging to acquire complete well-logging data along a borehole due to mechanical, environmental, and cost issues [11,12]. Therefore, it is necessary to generate the Vp and Vs to determine the presence of gas hydrate through BSR detection.

Synthesizing well-logging data in missing intervals using available petrophysical data through machine learning techniques has been investigated as a remedy, including the synthesis of acoustic logs [13–15]. Onalo et al. [13] adopted an artificial neural network (ANN) algorithm to estimate an acoustic log using shale volume, gamma-ray log (GR), and bulk density log (RHOB) as input variables in shale formation. Wang and Peng [14] generated synthetic vs. log data comparable to reference data using two algorithms: ANN-Levenberg–Marquardt and Extreme Learning Machine. Dalvand and Falahat [15] used an ANN to estimate the shear velocity using Vp, porosity (ϕ), RHOB, and GR. However, the conventional ANN is vulnerable in preserving the continuity of the sequential data, limiting the reliability of the prediction results [16,17].

More advanced neural networks have been recently deployed to preserve the data sequence in generating synthetic logging data. A recurrent neural network (RNN) is an algorithm with a recurrent flow where current state information is derived from the coaction between the inputs of the current step and outputs of the previous step [18]. Using an RNN-based algorithm, such as RNN, long short-term memory (LSTM), or a gated recurrent unit, is suitable for handling well-logging data because these data exhibit sequential characteristics similar to time-series data [19].

However, an RNN has limitations in maintaining old memory due to gradient vanishing [20]. LSTM is an algorithm that can resolve the shortcomings of RNN [20]. It has the advantage of extracting information from sequential data based on the RNN algorithm, preserving the impact of distant data in the long term [20]. Accordingly, LSTM has been conducted for operating well-logging data with improved accuracy [21–23]. Pham et al. [22] estimated Vp using an LSTM algorithm that inputted GR, RHOB, and ϕ from UK continental shelf wellbores to estimate acoustic logs, and then evaluated its performance against Gardner's equation. Zhang et al. [23] generated a synthetic vs. log from GR, RHOB, compressional travel time, neutron porosity (TNPH), photoelectric absorption (PEF), and resistivity (RT) using an LSTM algorithm with an accuracy of 98.9%.

Because compressional and shear logs are essential variables in assessing the presence of gas hydrate, studies have used these acoustic logs as inputs to estimate the petrophysical features of gas hydrate fields. Lee and Waite [24] estimated pore-space gas hydrate saturation using sonic velocities in gas-hydrate-bearing sands based on their relationships. Haines et al. [25] qualitatively evaluated gas hydrate saturation from the Vp and Vs of the Alaska North Slope Hydrate 01 wellbore using effective medium theory [26] and the Lee and Collett [27] approach. You et al. [28] predicted Vs using LSTM with GR, RT, RHOB, ϕ, and Vp as input variables for a gas hydrate field in the Gulf of Mexico. They claimed that the LSTM model with a coefficient of determination (R^2) above 0.85 was superior to the least-squared fitting model. In their work, the input variables were selected based on the input–output correlation analysis using the Pearson, Kendall, Spearman, and maximum correlations instead of conventional well-log variables (e.g., GR, RT, and RHOB).

However, acoustic logs are often not acquired, thus limiting accurate geomechanical interpretation and modeling [29]. Therefore, the capability to synthesize acoustic logs such as Vp is necessary to predict missing data, correct poor-quality data, manage reservoir uncertainty, and improve the interpretation of seismic attenuation coupled with well-logging data [30]. Karimpouli and Tahmasebi [31] estimated the solution of the seismic acoustic wave velocity based on the machine learning algorithm coupled with physics governing laws. Furthermore, more hydrate-field case studies are needed to demonstrate the synthesis capability considering that the number of hydrate fields is much smaller than the number of conventional hydrocarbon reservoirs worldwide.

The Ulleung Basin Gas Hydrate (UBGH) is being developed after BSR was detected in the southwestern part of the Ulleung Basin in 1996 [32]. Since 1996, UBGH-related projects and studies have been conducted by several research institutions, such as the Korea Institute of Geoscience and Mineral Resources, Korea Gas Corporation, and Korea National Oil Corporation [32]. At two UBGH wellbores (UBGH1-9 and UBGH1-10), Kim et al. [33] identified the BSR and estimated gas hydrate saturation using RT and Vp. Lee and

Collett [34] calculated gas hydrate saturation using the modified Biot–Gassmann Theory. Moridis et al. [35] evaluated the feasibility of gas production in the UBGH by estimating gas production at the UBGH2-6 wellbore. Lee et al. [36] conducted numerical simulations to estimate gas hydrate production at the UBGH through a cyclic depressurization method, which varies the depressurization stages based on bottom hole pressure changes. Park et al. [37] quantified the mineral composition of gas-hydrate-bearing sediments in the UBGH using machine learning techniques such as a convolutional neural network (CNN), RNN, ANN, and random forest (RF). Furthermore, the gas-hydrate-bearing zone in the UBGH is less consolidated; thus, its porosity range is larger than that of conventional reservoirs [27]. We intend to proceed with more case studies using hydrate field data to enhance UBGH development.

This study aims to generate synthetic Vp in distant wellbores using an LSTM model from easy-to-acquire well-logging variables (e.g., GR, RHOB, and RT), with application to the Ulleung Basin in the East Sea, Republic of Korea. This paper focuses on synthetic Vp generation due to the absence of Vs at the well-logging data of the UBGH field. First, conventional well-logging features are chosen as the input variables to design a learning-based predictive model. Second, ANN and LSTM models are designed based on the selected features. These models are trained to recognize geological relationships among the input and output well-logging features. Third, the performance of the models (ANN and LSTM) is compared to identify a suitable approach to handle well-logging data. Finally, k-fold cross-validation is conducted to validate the model's generality in the gas hydrate field.

2. Field Description

Figure 1 illustrates the locations of the UBGH wellbores in the Ulleung Basin in the East Sea, Republic of Korea [38]. Ulleung Basin is a back-arc basin that borders the Oki Bank, Korean Plateau, Japanese Arc, and Korean Peninsula [39]. The UBGH drilling activities are categorized into the first (UBGH1) and second drilling expeditions (UBGH2). The purposes of the two expeditions were the accurate evaluation of gas hydrate reserves and well placement for test production.

The reserve of gas hydrate in the UBGH was estimated as 620 million tons. The BSR was detected between 100 and 250 m below seafloor (m b.s.f.) [32,40]. This study focuses on analyzing the UBGH2 field with a production test wellbore. In UBGH2, gas hydrate was found within thin sand layers in mud. The thickness of the sand was estimated to be less than 50 cm [35,41].

As depicted in Figure 1, the UBHG2 drilling expedition includes 13 wellbores: UBGH2-1_1, UBGH2-1_2, UBGH2-2_1, UBGH2-2_2, UBGH2-3, UBGH2-4, UBGH2-5, UBGH2-6, UBGH2-7, UBGH2-8, UBGH2-9, UBGH2-10, and UBGH2-11 [38]. Based on logging while drilling or measurement while drilling, their borehole data were acquired as follows: Vp, GR, RHOB, PEF, RT (shallow/medium/deep), TNPH, caliper, bulk density correlation, equivalent circulating density, downhole annulus temperature, downhole annulus pressure, collar rotational speed, rate of penetration averaged over the last 5 ft, density time after bit, neutron time after bit, resistivity time after bit, and sonic time after bit. The logging data were measured at every 0.1524 m (0.5 ft) interval.

Figure 2 is the distance matrix of the UBHG2 wellbores. These distances were calculated based on the longitude and latitude data of the wellbores. The closer the inter-well distance, the more similar the geologic features between the wellbores [42]. The average inter-well distance is 65.53 km. Wellbores with similar names, such as UBGH2-2_1 and UBGH2-2_2, are in close proximity. Wellbore UBGH2-3 is the most distant wellbore among the UBGH2 wellbores, with an average distance of 105.08 km.

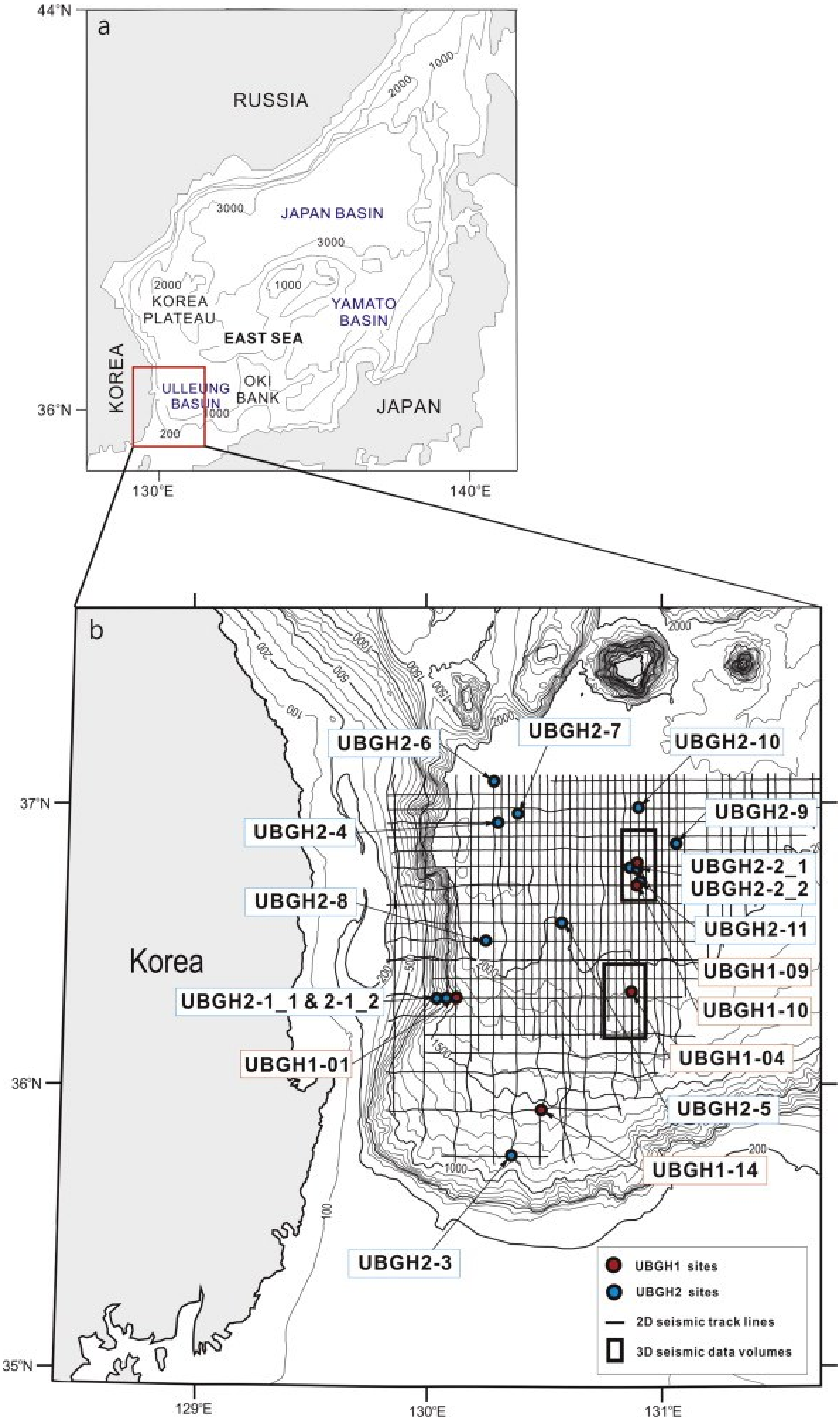


Figure 1. (**a**,**b**) Location of the UBGH boreholes in Ulleung Basin, Republic of Korea [38] (Reprinted with permission from Ref. [38]).

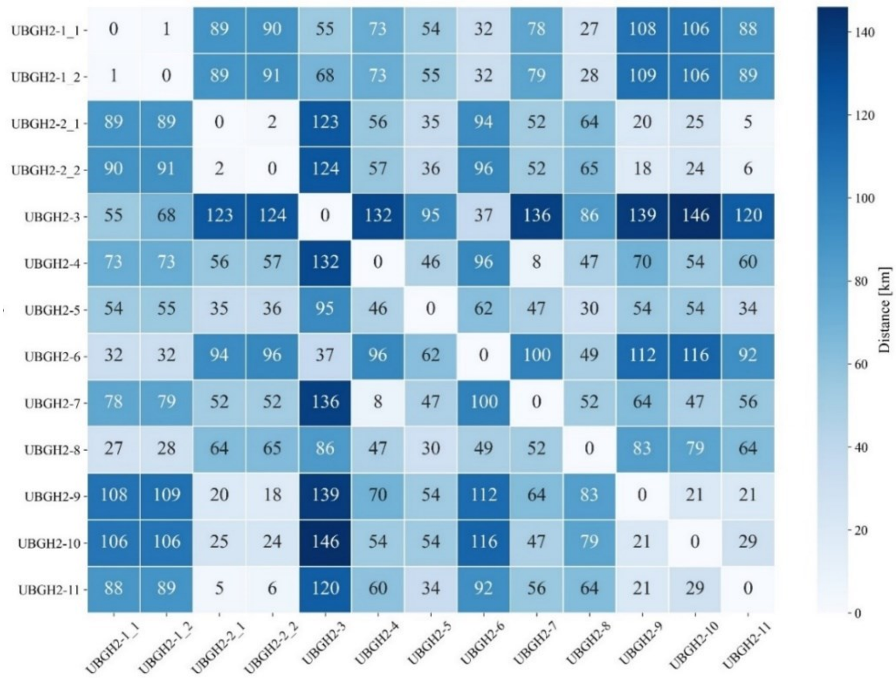


Figure 2. Distance matrix between 13 wellbores in the UBGH2 field.

Based on core data, the estimated porosity in the UBGH2 field ranges from 26.77 to 94.07%. In-depth studies were conducted for 6 of the 13 UBGH2 wellbores (UBGH2-2-1, UBGH2-2-2, UBGH2-6, UBGH2-9, UBGH2-10, and UBGH2-11), which had relatively large gas hydrate resources, with a porosity range between 63.96% and 71.35% [43,44]. Based on seismic survey results, UBGH2-6 was selected and operated as a production test well because of its production potential [45]. This borehole had the thickest hydrate-bearing deposit among all UBGH wellbores [46]. Its water depth was 2153 m, and gas hydrates were deposited between 140 and 154 m b.s.f.

3. Methodology

3.1. Artificial Neural Network (ANN)

An ANN is a fundamental machine learning algorithm based on the neurons' data processing [47,48]. It identifies patterns among the input and output factors, extracts feature information from data, and establishes a nonlinear relationship between inputs and outputs [47,48]. Figure 3 illustrates the structure of a feedforward ANN. A conventional ANN consists of an input layer to import the raw data, one or more hidden layers where the features are extracted, and an output layer to derive the output results. Nodes in a layer are connected to those in the next layer with weights.

During feedforward, inputs (x) are multiplied with weights (ω) and added to a bias (b), and then applied to the activation function (σ). Consequently, outputs (y) are calculated, as depicted in Equation (1). This process is repeated from the input layer to the output layer via the hidden layer(s). The ANN is trained until the weights are optimized through backpropagation [47,48]. Thus, the ANN performance depends on its structure (the number of hidden layers and nodes in each layer).

$$y = \sigma(\omega x + b). \tag{1}$$

3.2. Long Short-Term Memory (LSTM)

Figure 4 illustrates the LSTM structure [19,49,50]. A unit cell of LSTM receives, computes, and transfers a cell state and a hidden state. The cell state refers to a status of filtered data flow through the cell, considered long-term memory, whereas the hidden state is considered short-term memory [19,49,50].

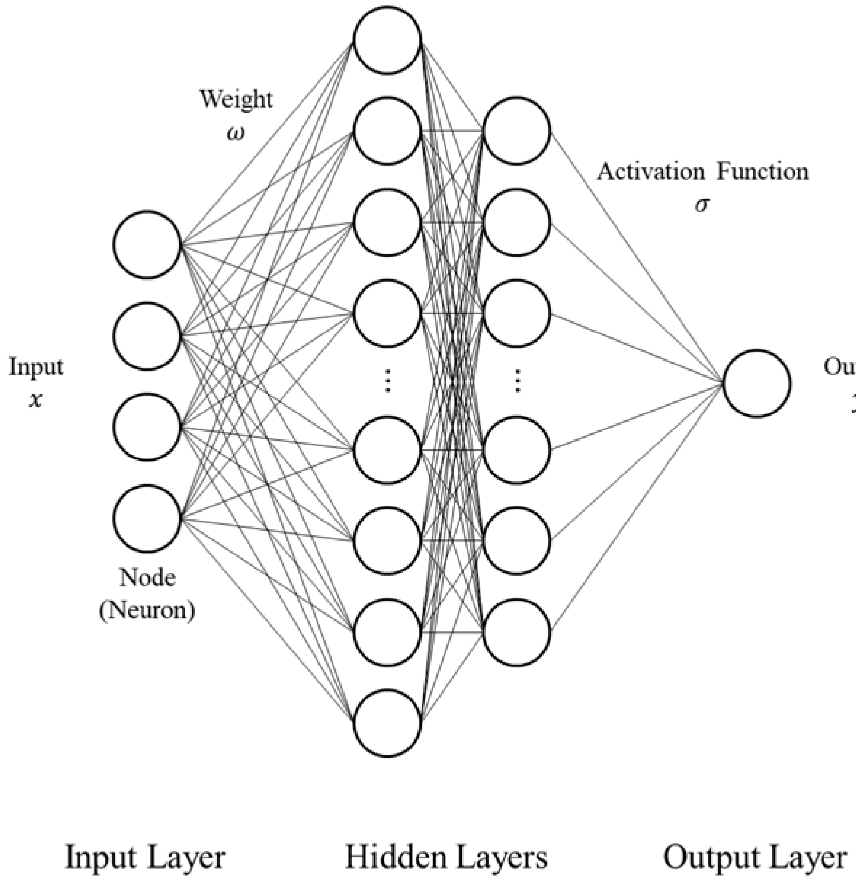


Figure 3. ANN structure.

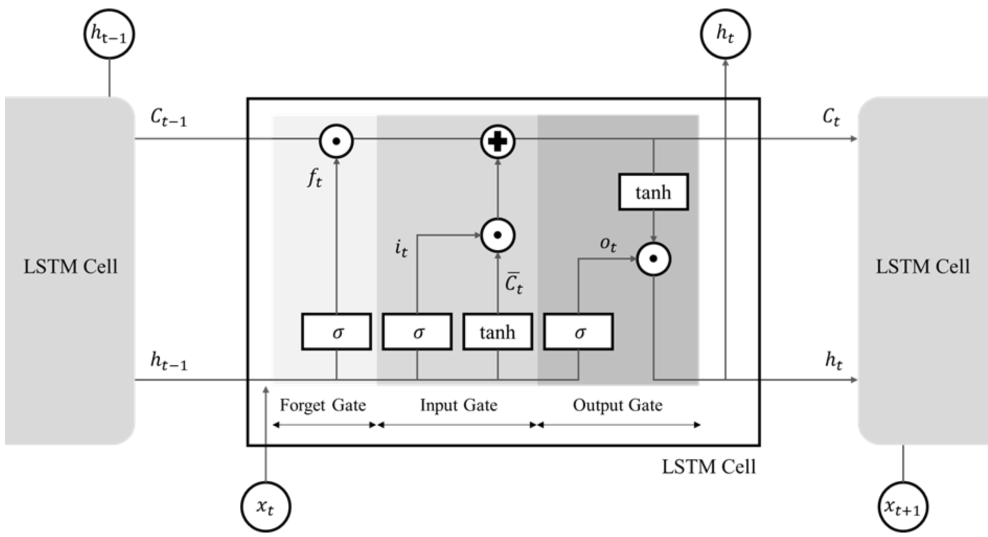


Figure 4. LSTM structure with a unit cell.

The unit cell consists of three gates to control data flow: the forget gate, input gate, and output gate. Let f_t, i_t, o_t, and h_t be the outputs of the forget gate, input gate, output gate, and tanh layer at the current t-th step, respectively [19,49,50]. Furthermore, the subscripts f, i, o, and h are the forget gate, input gate, output gate, and tanh layer, respectively.

The forget gate f_t conducts selective preservation of data transferred from the previous hidden state h_{t-1} and current input data x_t using Equation (2):

$$f_t = \sigma(\omega_f \cdot [h_{t-1},\ x_t] + b_f), \tag{2}$$

where ω is the weight, b is the bias, and σ is the sigmoid function. As σ converges to zero, data are forgotten, while σ = 1 indicates complete data preservation [19,49,50].

The input gate i_t determines whether to store new data through two stages. First, new data in i_t are filtered from h_{t-1} and x_t (Equation (3)). Second, the hyperbolic tangent layer (i.e., tanh layer) generates candidate values $\overline{C_t}$ that can be updated in the cell state (Equation (4)):

$$i_t = \sigma(\omega_i \cdot [h_{t-1},\ x_t] + b_i), \tag{3}$$

$$\overline{C_t} = \tanh(\omega_c \cdot [h_{t-1},\ x_t] + b_c). \tag{4}$$

Here, $\overline{C_t}$ is the candidate value and tanh is the hyperbolic tangent function.

The cell state receives data selectively from i_t and $\overline{C_t}$. The current cell state C_t is updated using Equation (5), where $\bullet$ is the Hadamard product:

$$C_t = f_t \bullet C_{t-1} + i_t \bullet \overline{C_t}. \tag{5}$$

Finally, the unit cell exports the current hidden state h_t using Equations (6) and (7). The amount of data from h_{t-1} and x_t to be released is determined in the output gate o_t (Equation (6)). Then, h_t is derived through Equation (7).

$$o_t = \sigma(\omega_o \cdot [h_{t-1},\ x_t] + b_o), \tag{6}$$

$$h_t = o_t \bullet \tanh(C_t). \tag{7}$$

3.3. Data Pre-Processing

This study aimed to design a versatile model to predict Vp based on conventional well-logging data. Accordingly, the optimal input combination is searched considering GR, RHOB, RT, PEF, and TNP among 17 logging data types addressed in Section 2. The scope of this study is synthesizing Vp from 95 to 255 m b.s.f. to include the BSR range between 100 and 250 m b.s.f. Because the logging data interval was 0.5 ft, the number of logging data points was 1050 for every wellbore except for UBGH2-11, with 878 logging data points—its well depth (228.78 m b.s.f.) being smaller than the upper limit of this analysis (255 m b.s.f.).

A robust scaler was used to minimize the effects of outliers on the overall training performance of a neural network. For every logging data type, logging data were normalized using the robust scaler:

$$x' = \frac{x - Q_2}{Q_3 - Q_1}, \tag{8}$$

where x and x' are the original datum and corresponding scaled datum at an arbitrary measurement point, respectively. Q_1, Q_2, and Q_3 are the first, second, and third quartiles of the logging data.

Before determining the input combination, we calculated the coefficient of determination (R^2) between each input logging variable and the output logging variable (Vp) for all wellbores (Table 1). On average, the output Vp exhibited a high correlation in the descending order of TNPH, RHOB, RT, GR, and PEF. The difference between the minimum and maximum was greater than 0.3, yielding a significant variance for every input logging variable. Because it was insufficient to match the input and output logging variables one by one, the neural network model was used to capture the nonlinear relationships between multiple inputs and output.

Table 1. Coefficient of determination between the input variable and Vp in each wellbore.

Wellbore Name	Input Variable				
	GR	RT	RHOB	TNPH	PEF
UBGH2-1_1	0.00004	0.00023	0.27144	0.21068	0.14138
UBGH2-1_2	0.15054	0.33640	0.01392	0.02958	0.10049
UBGH2-2_1	0.47748	0.02856	0.25503	0.52418	0.00063
UBGH2-2_2	0.11022	0.07453	0.44090	0.47472	0.06503
UBGH2-3	0.25000	0.36120	0.32149	0.29594	0.26214
UBGH2-4	0.02789	0.01000	0.37577	0.32604	0.05856
UBGH2-5	0.00518	0.41474	0.66912	0.64160	0.27040
UBGH2-6	0.17140	0.08352	0.31584	0.29052	0.01103
UBGH2-7	0.24404	0.33989	0.30250	0.27458	0.22090
UBGH2-8	0.00012	0.19536	0.55801	0.41732	0.31923
UBGH2-9	0.02690	0.16241	0.18490	0.22468	0.01210
UBGH2-10	0.06812	0.10758	0.06350	0.12110	0.00130
UBGH2-11	0.05063	0.22753	0.15524	0.12180	0.04040
Average	0.12173	0.18015	0.30213	0.30406	0.11566
Minimum	0.00004	0.00023	0.01392	0.02958	0.00063
Maximum	0.47748	0.41474	0.66912	0.64160	0.31923
Standard Deviation	0.13854	0.14451	0.18346	0.17392	0.11485

3.4. Generation of Synthetic Vp Log

Figure 5 describes the procedure to generate synthetic Vp in this study. The first step is data pre-processing. Well-logging data are acquired from a target field, to determine and analyze the logging interval; subsequently, the well-logging data were normalized using a robust scaler (Figure 5a). The second step is selecting the input logging variables (Figure 5b). Input selection is followed by the third step—designing the structure of the deep-learning-based predictive model (e.g., the number of hidden layers, activation function, and dropout rate) (Figure 5c). We then performed a hyperparameter sensitivity analysis to refine the predictive model and confirm the generality (Figure 5d). The process from Figure 5b to 5d was repeated for all available combinations of the input variables to search for the most efficient model. The efficiency is assessed in terms of two performance indicators: R^2 and root mean square error (RMSE) (Figure 5e):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}, \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{10}$$

where n is the number of data points, y_i is the reference value, $\hat{y}_i$ is the predicted value, and $\overline{y}$ is the mean value. Finally, k-fold cross-validation was conducted to verify model consistency (Figure 5f).

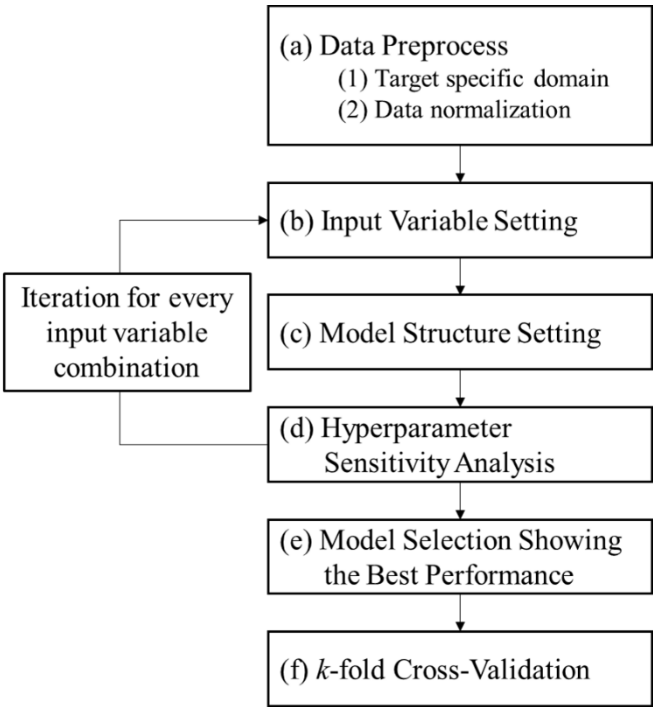


Figure 5. Flowchart for generating synthetic Vp.

4. Results and Discussion

Various studies generating synthetic Vp selected input variables based on feature-selection methods (e.g., RF and Pearson correlation) [28,51,52]. Although these methods select input variables highly correlated with Vp, they discard the characteristics inherent in each factor [53]. Therefore, a deep-learning model generating synthetic Vp with the highest performance is designed and performed in this section.

4.1. Structure of the LSTM Model

Table 2 summarizes the structure of the LSTM model used to generate the synthetic Vp. The LSTM model was composed of one input layer, one hidden layer, and one output layer, to simplify the learning-based model given the training data size. Adam (Adaptive Moment Estimation) was adopted as a neural network optimizer. The dropout rate for preventing overfitting was 0.25 [54]. Total data were categorized into training, validation, and test datasets: data from nine wellbores in the training dataset (70%), data from two wellbores in the validation dataset (15%), and data from two wellbores in the test dataset (15%). The loss function for model training was the mean square error (MSE). Early stopping was activated to prevent overfitting if the loss function stagnated for 20 epochs.

Table 2. Parameters of the LSTM model used to generate synthetic Vp.

Parameter	Value
Neural Network Algorithm	LSTM
Number of Layers (Input, Hidden, Output)	(1, 1, 1)
Optimizer	Adam
Dropout Rate	0.25
Ratio of Data (Training, Validation, Test)	(70%, 15%, 15%)

4.2. Performance Evaluation of the LSTM Model

This case study examined the robustness of the LSTM model in the synthetic Vp generation with sensitivity analysis on two parameters: the combination of input logging variables and the number of hidden neurons in the model. The model performance was assessed in terms of the R^2 and RMSE using Equations (9) and (10), respectively. For finding the optimal input combination among the input set (GR, RHOB, RT, PEF, and TNPH), up to five inputs were imported into the LSTM model. Because the number of hidden neurons affects the model performance [55], we set up the number of hidden neurons in powers of two (i.e., 2 m) and increased the exponent m from zero to seven for each learning-based model. Considering eight scenarios for the number of hidden neurons for each input combination, 248 (= 8 × 31) experiments were conducted in total, where 31 = $_5C_1$ experiments from one input (Table A1) + $_5C_2$ experiments from two inputs (Table A2) + $_5C_3$ experiments from three inputs (Table A3) + $_5C_4$ experiments from four inputs (Table A4) + $_5C_5$ experiment from five inputs (Table A5). Refer to Appendix A for the performance evaluation results of all experiments.

Figure 6 illustrates radar charts to compare the R^2 values obtained using the LSTM model for the 248 experiments: training results (Figure 6a), validation results (Figure 6b), and test results (Figure 6c). Nine wellbores, UBGH2-1_1, UBGH2-1_2, UBGH2-2_1, UBGH2-3, UBGH2-4, UBGH2-5, UBGH2-7, UBGH2-8, and UBGH2-9, were used for training, UBGH2-10 and UBGH2-11 for validation, and UBGH2-2_2 and UBGH2-6 for testing. UBGH2-6 and UBGH2-2_2 were selected as test data to judge whether the LSTM model could generate synthetic Vp with reliability. We intended to include UBGH2-6 in the LSTM test dataset because it was the only production test well in the UBGH field [32]. Similar to UBGH2-6, UBGH2-2-2 was a wellbore drilled at high-quality gas-hydrate-bearing sediments [44].

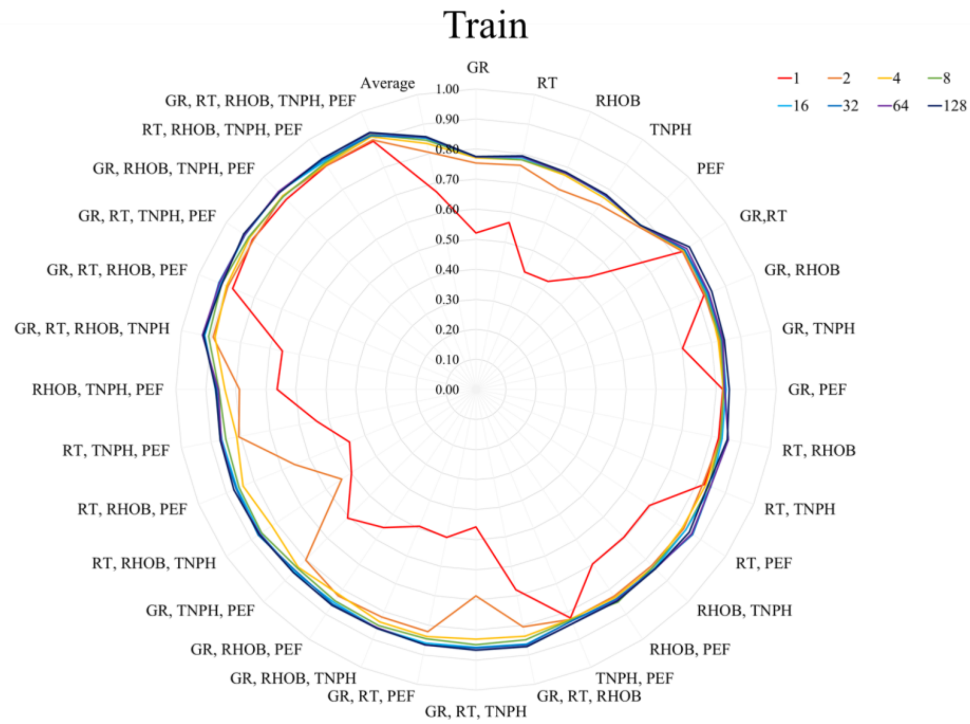
For the experiments with a single input variable (Table A1), the average R^2 for the training, validation, and test data were 0.720, 0.446, and 0.514, respectively. In most experiments, the training R^2 values were above 0.70 when using more than two neurons in the hidden layer. However, the average R^2 values for validation and test data were approximately 0.50, which is less satisfactory than the training results. The model obtained high performance in the descending order of inputting PEF, TNPH, RHOB, GR, and RT, with average R^2 values of 0.593, 0.581, 0.573, 0.571, and 0.553, respectively. These values improved by more than the Pearson correlation coefficients based on linearity in Table 1. The sensitivity analysis results for a single input variable indicate that LSTM is more efficient in capturing a nonlinear relationship between input and output.

The combination with four input variables had the highest average R^2 value among the five input combinations. As illustrated in Table A4, the highest performance was attained from the experiment inputting GR, RHOB, TNPH, and PEF with 64 hidden neurons: R^2 values of 0.930, 0.711, and 0.8481 for the training, validation, and test datasets, respectively.

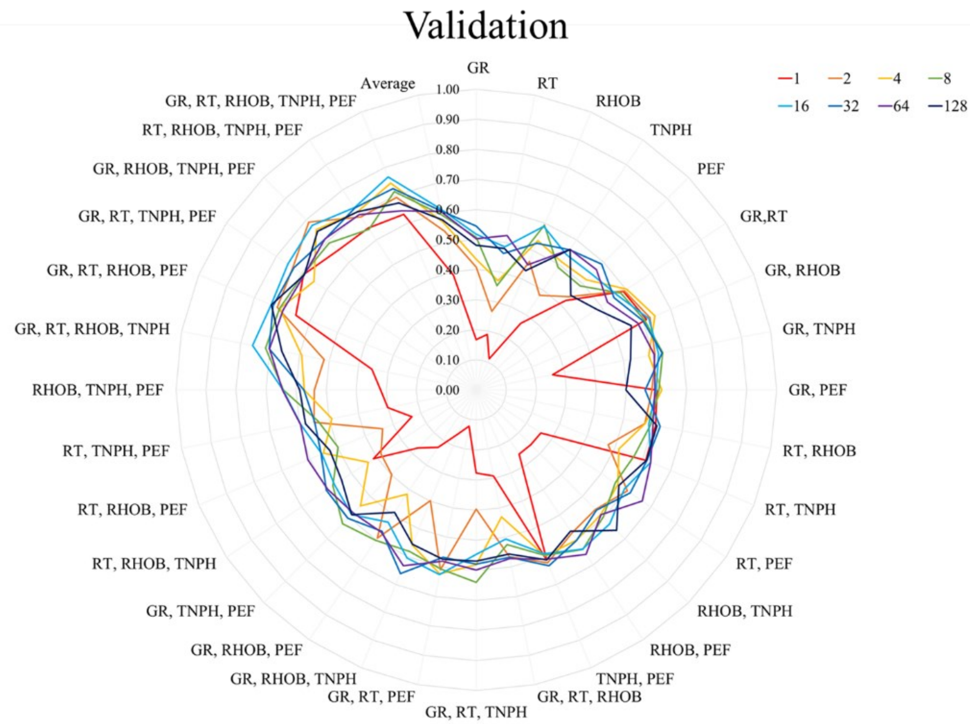
Furthermore, it is essential to balance accuracy and computational cost. The LSTM model with 64 hidden neurons was superior to the model with 128, while the latter required a more complex neural network structure. The numbers of trainable LSTM parameters were 17,729 and 67,713, with 64 and 128 hidden neurons, respectively. Based on these results on accuracy and computational cost, we decided to use the LSTM models with 64 hidden neurons. Moreover, the input set of GR, RHOB, TNPH, and PEF was selected as the optimal input combination for the model for subsequent analysis.

An ANN model was built to compare its performance with that of the LSTM model to confirm the suitability of the LSTM model for sequential logging data analysis. Under the same data conditions and neural network configuration, we designed the structure of the ANN model similar to that of the LSTM model in terms of the numbers of trainable parameters: 17,813 for the ANN model and 17,729 for the LSTM model.

Figure 7 illustrates the scatter plots between the reference and generated Vp data obtained using the two neural network models. This case corresponds to Fold 1 of the 10-fold cross-validation. R^2 values for the training, validation, and test data obtained using the LSTM model were 0.931, 0.757, and 0.851, respectively. In contrast, the ANN model yielded unsatisfactory values of 0.415, 0.007, and 0.395. Furthermore, the RMSE decreased by more than 0.3 when applying the LSTM model compared with the ANN model. The ANN model failed in generating synthetic Vp.

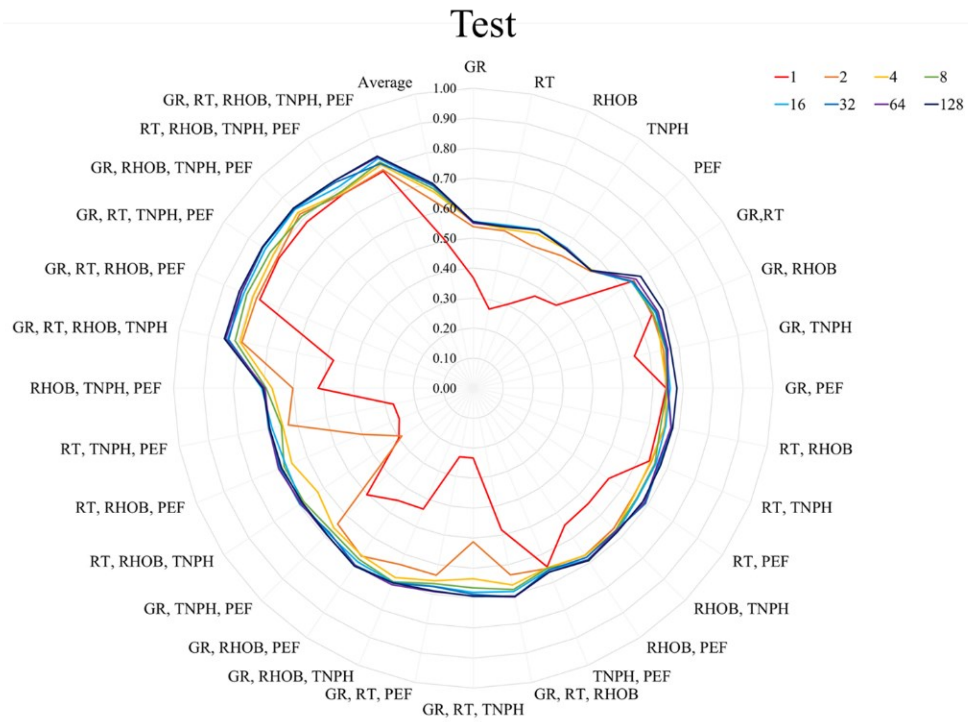


(**a**) Training data



(**b**) Validation data

Figure 6. *Cont.*



(**c**) Test data

Figure 6. Comparison of the R^2 values for input combinations at various numbers of hidden neurons.

Interestingly, the validation performance also deteriorated for the LSTM model when the reference Vp was larger than 6 because of incompatibility between the training and validation data ranges. Training and test data mostly ranged from -2 to 6 and validation data from -2 to 12. Furthermore, the training data were sparsely distributed when the reference Vp was larger than -6. These results imply the intrinsic limitation of the machine learning advantage in interpolation compared to extrapolation.

The influence of the data range was investigated by re-comparing the performance of the two neural network models with different training and validation data (Figure 8); this case is Fold 3 of the 10-fold cross-validation. Wellbores UBGH2-10 and UBGH2-11 were included in the training data, and wellbores UBGH2-1_1 and UBGH2-1_2 were included in the validation data. Switching these data produced similar ranges for training and validation, improving the validation performance for both models while maintaining their training and test performance: The R^2 of ANN increased from 0.007 to 0.119, and the R^2 of LSTM increased from 0.757 to 0.920. This result highlights the importance to let each dataset (training, validation, and test) have a similar data distribution for making an accurate machine learning model.

Nevertheless, the ANN's poor R^2 values reveal a vulnerability in the sequential data analysis. Furthermore, despite the validation performance, the test performance of the LSTM model was satisfactory in each figure because the range of the test data was within the training data. Based on these results, we attest that the LSTM model outperforms the ANN model for sequential data analysis and reconfirm the significance of balancing the training and validation data ranges in deep learning.

We also conducted a sensitivity analysis for the learning rate and batch size to create a cost-effective LSTM model. Determining the proper values for these hyperparameters is vital because a large learning rate might cause overshooting, while a small learning rate requires expensive computational costs [56]. Batch size also affects the model generality [57]. The effects of batch size and learning rate on R^2 and computational time were analyzed for test data using the LSTM model depicted in Figure 7. The computational time was the arithmetic mean from four learning rate cases in each batch size.

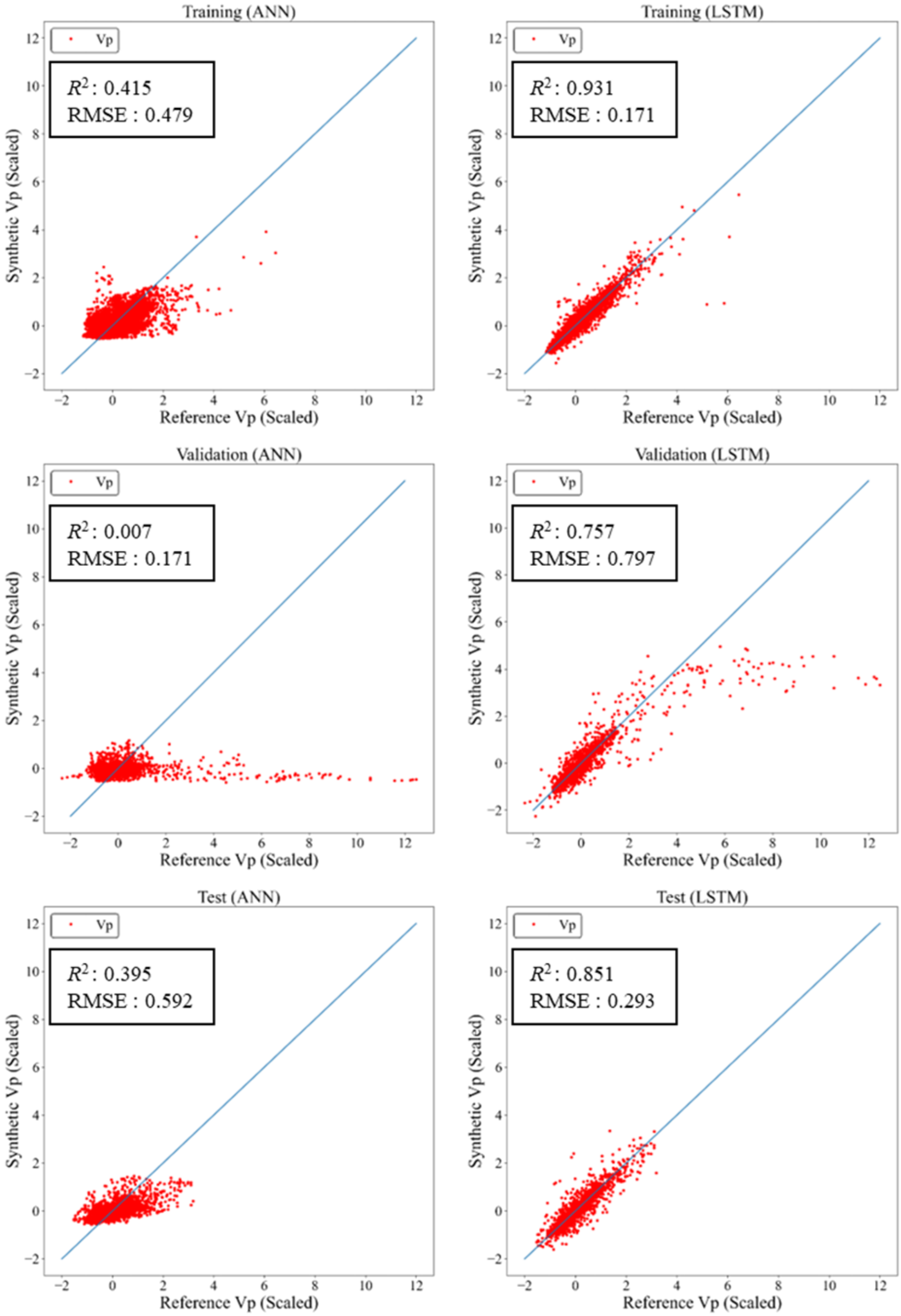


Figure 7. Scatter plots comparing the reference and predicted Vp data (scaled) for the training, validation, and test data for the ANN and LSTM (Fold 1).

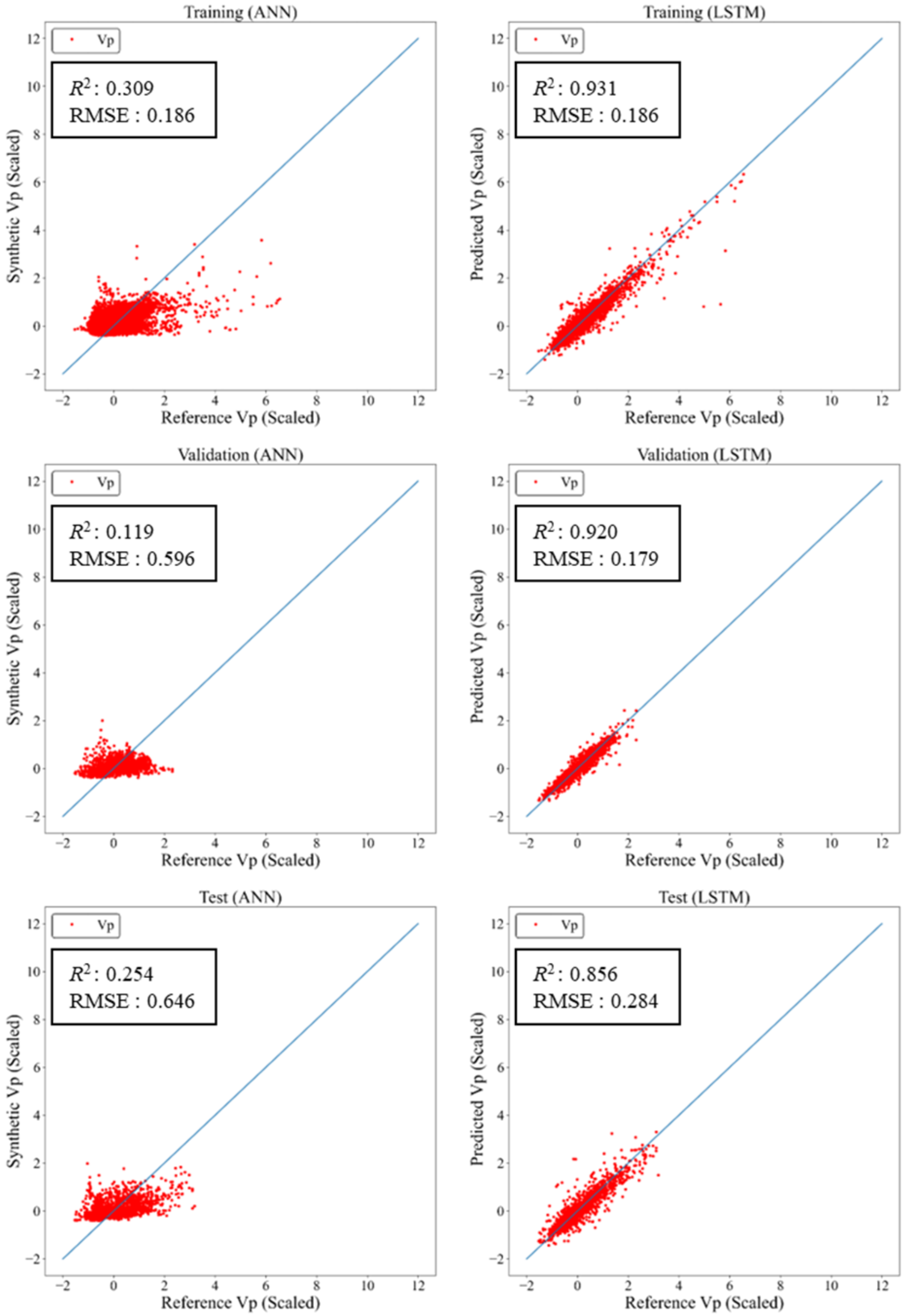


Figure 8. Scatter plots comparing the reference and predicted Vp data (scaled) for the training, validation, and test data for the ANN and LSTM (Fold 3).

Figure 9 confirms that the model performance depends on the two hyperparameters—the larger the batch size, the lower the computational cost as R^2 decreases. When the learning rate was 10^{-2}, the performance increased as the batch size increased up to 32.

The performance with a learning rate of 10^{-5} was relatively inferior. A learning rate of 10^{-3} and batch size of 32 had the highest R^2 at an affordable computational cost. With these hyperparameter values, the R^2 of each experiment was greater than 0.75, confirming both the consistency and generality of the LSTM model on well-logging data in the gas hydrate field. Thus, this hyperparameter setting was used for all experiments throughout this study.

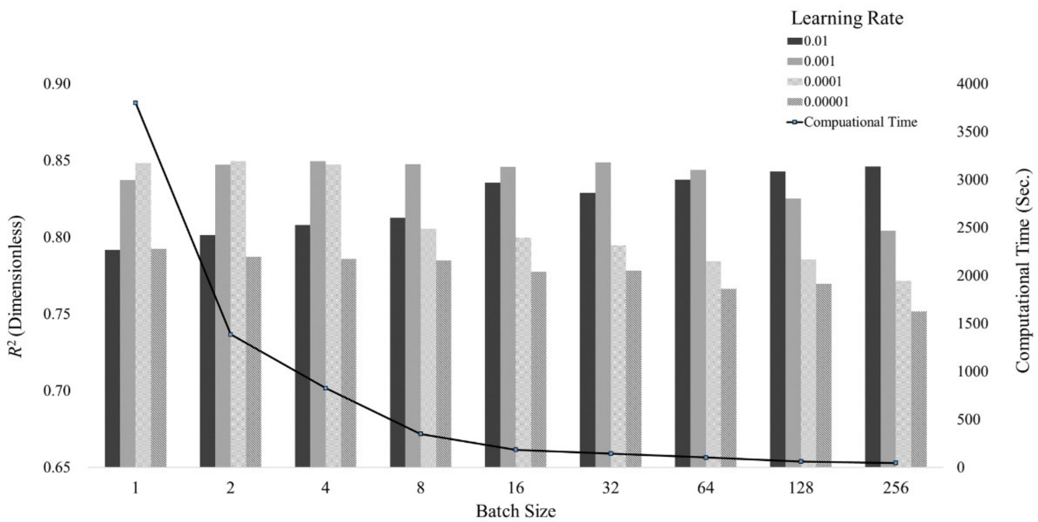


Figure 9. Sensitivity analysis results on the learning rate and batch size for the LSTM model.

4.3. k-Fold Cross-Validation

k-Fold cross-validation was conducted to verify the LSTM model's consistency [50]. For 13 wellbores, *k* was set to 10, which is typical for cross-validation. Figure 10 depicts the composition of the training, validation, and test datasets. Each fold was composed of training data from nine wellbores, validation data from two wellbores, and test data from two wellbores. Wellbores UBGH2-2_2 and UBGH2-6 were fixed as the test data for a fair comparison of the test performance according to the training and validation performance variation.

Figure 11 compares the performance of the LSTM model for each fold of the 10-fold cross-validation. The performance of each fold was assessed in terms of R^2 and RMSE. The model's consistency was quantified in terms of two statistical parameters: the average (μ) and standard deviation (σ) of each indicator (e.g., $\mu \pm \sigma$). The performance of every experiment was satisfactory for generating synthetic Vp using the LSTM model. The average $\mu \pm \sigma$ of the R^2 values were 0.932 ± 0.006 for the training data, 0.872 ± 0.091 for the validation data, and 0.853 ± 0.002 for the test data. The average $\mu \pm \sigma$ of the RMSE were 0.200 ± 0.277 for the training data, 0.322 ± 0.254 for the validation data, and 0.288 ± 0.003 for the test data.

For the training data, R^2 ranged from 0.922 to 0.943. For the validation data, R^2 ranged from 0.673 to 0.959. For the test data, R^2 ranged from 0.851 to 0.856. For the test data, Fold 3 had the highest performance among the 10 folds. Its R^2 values were 0.931, 0.920, and 0.856 for the training, validation, and test data, respectively.

Table 3 summarizes the ranges of the scaled logging data in each fold. The values were calculated using the robust scaler (Equation (8)). A high R^2 value accompanied a low RMSE. The training data range covered that of the validation data except for Folds 1 and 2. These two folds had larger RMSEs and smaller R^2 values than the other folds because the range of validation data exceeded that of the training data, as indicated in Figures 7 and 8.

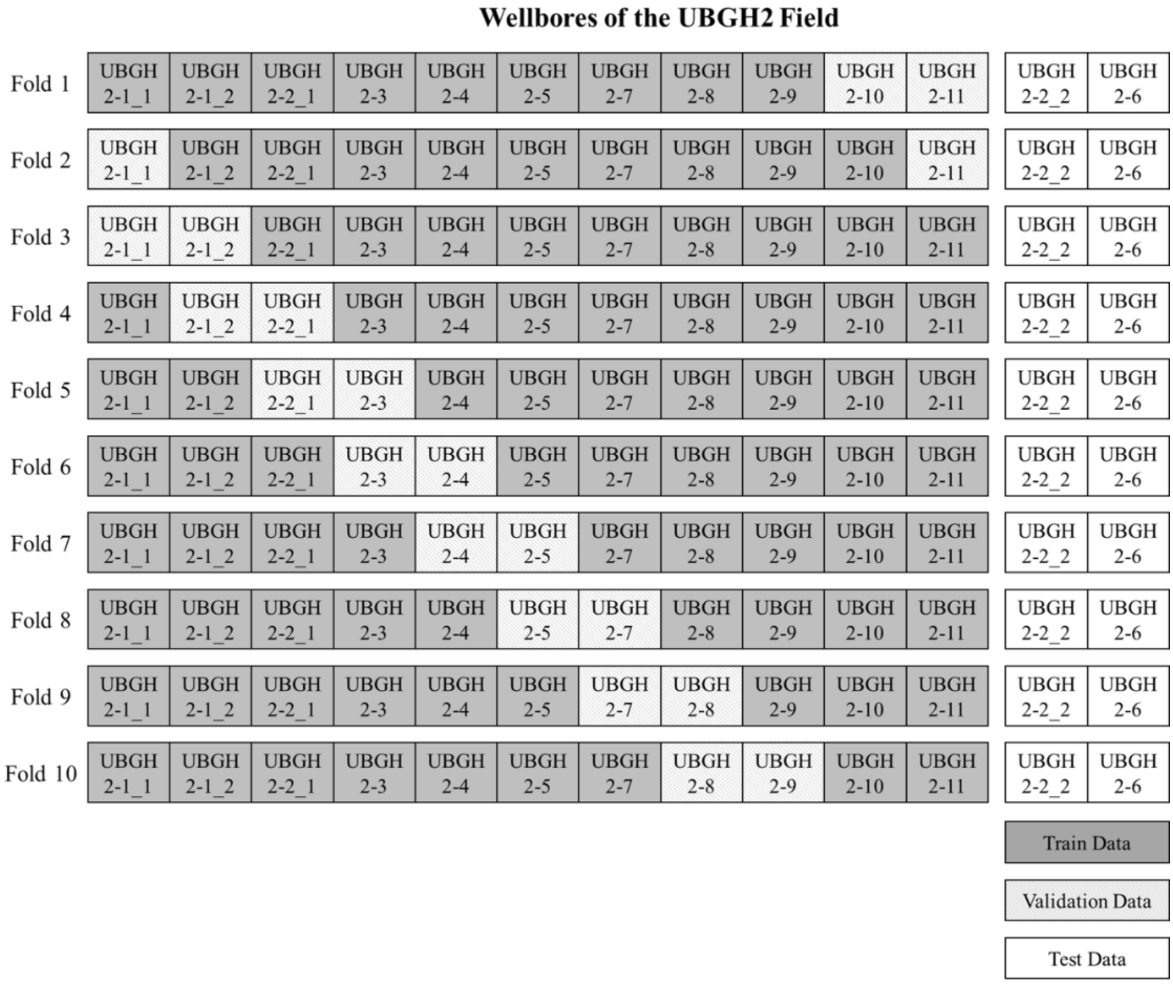


Figure 10. Training, validation, and test datasets in each fold of the 10-fold cross-validation for Vp prediction.

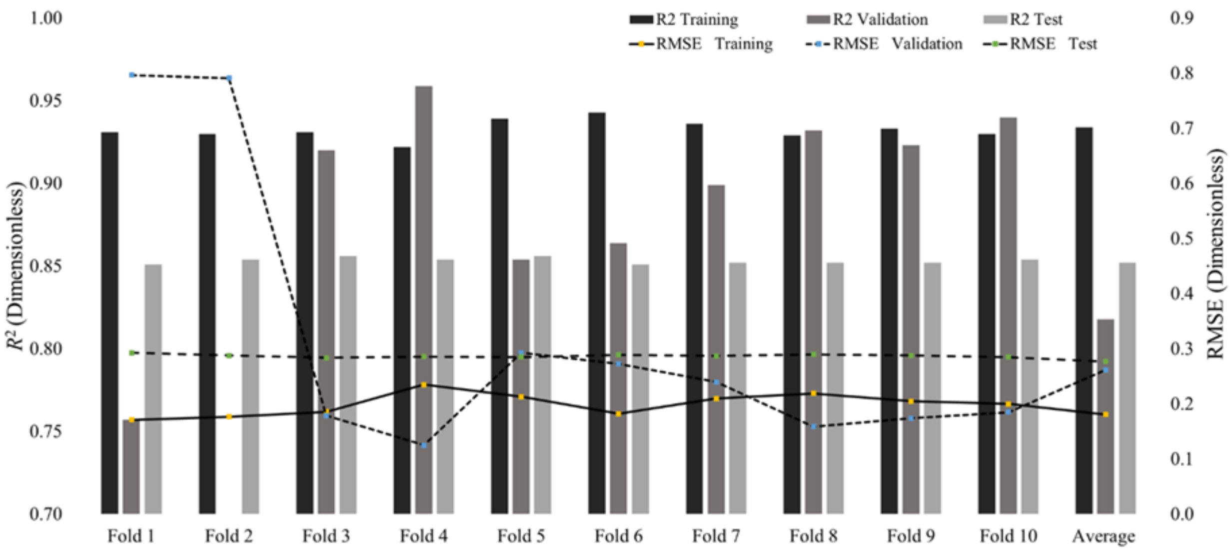


Figure 11. Performance of the 10-fold cross-validation obtained using the LSTM model.

Table 3. Range of scaled data in each fold of the 10-fold cross-validation.

Fold Number	Data Type	Mean	Standard Deviation	Minimum	Maximum
Fold 1	Training	0.0870	0.6500	−1.1556	6.4422
	Validation	0.2624	1.4909	−2.3569	12.4731
	Test	0.0071	0.7469	−1.5295	3.1866
Fold 2	Training	0.1042	0.6668	−1.083	6.4581
	Validation	0.2731	1.3162	−2.1941	12.2122
	Test	0.0071	0.7469	−1.5295	3.1866
Fold 3	Training	0.1228	0.6995	−1.5363	6.5456
	Validation	0.0054	0.6352	−1.5293	2.3126
	Test	0.0071	0.7469	−1.5295	3.1866
Fold 4	Training	1.6254	0.8436	−1.7872	8.3321
	Validation	0.0488	0.6029	−1.1615	3.9717
	Test	0.0071	0.7469	−1.5295	3.1866
Fold 5	Training	0.1475	0.8598	−1.9181	9.0759
	Validation	−0.1290	0.7534	−1.5396	6.1157
	Test	0.0071	0.7469	−1.5295	3.1866
Fold 6	Training	0.1372	0.7552	−0.4369	7.6077
	Validation	0.1283	0.7686	−1.1311	6.9952
	Test	0.0071	0.7469	−1.5295	3.1866
Fold 7	Training	−0.1565	0.8117	−1.7292	8.1865
	Validation	0.0776	0.7495	−1.4006	3.3819
	Test	0.0071	0.7469	−1.5295	3.1866
Fold 8	Training	0.1349	0.8198	−1.8158	8.3387
	Validation	0.0327	0.6069	−0.9845	2.5030
	Test	0.0071	0.7469	−1.5295	3.1866
Fold 9	Training	0.1435	0.7756	−1.6429	7.5061
	Validation	0.0513	0.6240	−0.9654	2.1172
	Test	0.0071	0.7469	−1.5295	3.1866
Fold 10	Training	0.1328	0.7367	−1.6206	7.0720
	Validation	0.1702	0.7270	−1.1042	2.9059
	Test	0.0071	0.7469	−1.5295	3.1866

In contrast, Folds 3 to 10 obtained R^2 values greater than 0.85 for both the training and validation data because the training range included the validation range. Folds 1 and 2 yielded R^2 values of 0.812 and 0.789 for validation. Although the coincidence in the data range is desirable, the acceptable performance from the cross-validation implies the robustness of the proposed approach for generating synthetic logs using LSTM to estimate parameters in distant wellbores.

Figure 12 illustrates the synthetic Vp logs at the training, validation, and test wellbores against their reference data for Fold 1. All the synthetic and reference data were expressed in scaled values using Equation (8). Among the nine training wellbores, UBGH2-5 obtained the highest R^2 of 0.938, while UBGH2-4 obtained the lowest R^2 of 0.854. The overall trend of our estimation was comparable to the reference except for the mismatch at the depth near 140 m b.s.f. in the validation data because the scaled Vp values rapidly decreased in both validation wellbores UBGH2-10 and UBGH2-11, as already captured in Figures 6 and 7. The mismatch was resolved by data swapping, as depicted in Figure 8. We confirmed that the synthetic Vp was generated at a high quality based on the two test wellbores. Therefore, the designed LSTM model and selected input variables were suitable for generating synthetic Vp with less-consolidated gas-hydrate-bearing sediments in the UBGH2 field.

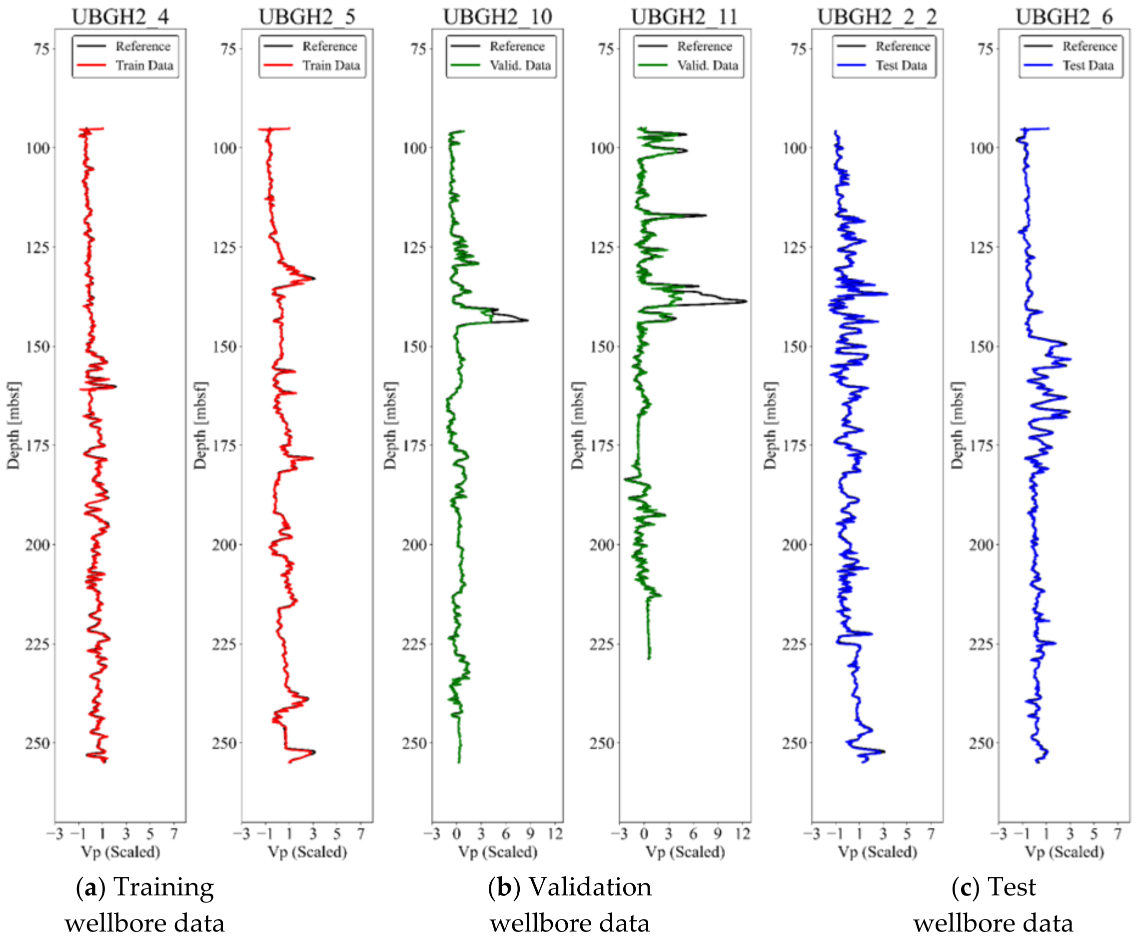


Figure 12. Comparison of synthetic and reference Vp profiles in the UBGH2 field: (**a**) training wellbore data (red curves), (**b**) validation wellbore data (green curves), and (**c**) test wellbore data (blue curves).

4.4. Discussion

The UBGH2 field case study validated that the LSTM model could generate synthetic log data in distant wellbores. Input variables were selected primarily based on the correlation with the output variable Vp. For the UBGH2 wellbores, RHOB had the highest correlation coefficient while GR had the lowest, as summarized in Table 1. Nonetheless, the Vp-learning model obtained the highest performance among the Vp-estimation results with a single input variable by capturing the nonlinear relationship between GR and Vp. Furthermore, the highest performance was derived using GR, RHOB, TNPH, and PEF as input variables.

The use of all five logging types by adding RT in the input set yielded a similar performance. The R^2 of using all five logging types was 0.3 less than the best performance on average. Thus, Figures 6–8 illustrate that using all given input variables does not guarantee improved performance for the learning-based predictive model. Furthermore, our results imply that the LSTM is suitable for capturing nonlinearity among the inputs and output. Except for wellbores with similar names (e.g., UBGH2-1_1 and UBGH2-1_2), Fold 1 and the other folds had R^2 values higher than 0.80. The influences of similar data distribution of training, validation, and test datasets were analyzed through Figures 7 and 8. The more similar the data distribution, the more accurate performances were obtained from the model. Therefore, the data distribution affects the overall performance of the designed model.

Due to the unavailability of core Vp data, this study focused on the synthetic generation of Vp logs. If available, the proposed LSTM approach could be used to integrate core and well-logging data. Our future research will use a learning-based model to generate high-resolution logging data compatible with core data.

5. Conclusions

This study developed an LSTM-based deep-learning model to generate synthetic Vp logs of distant wellbores in the gas-hydrate-bearing-sediments in the UBGH2 field in the Republic of Korea. Sensitivity analysis results of the input combination and hyperparameters (i.e., the learning rate and batch size) produced the optimal model structure, with accuracy and generality. Inputting GR, RHOB, TNPH, and PEF logs efficiently synthesized Vp logs, and satisfactory performance was achieved in terms of the RMSE and R^2 for 13 wellbores in the UBGH2 field. Hyperparameter analysis balanced the model's accuracy and computational cost.

The model's generality was also examined using 10-fold cross-validation because each fold yielded an R^2 higher than 0.8 on average. Data swapping between the training and validation demonstrated consistent test performance. The LSTM and ANN results comparison indicated that the LSTM-based model was more suitable than ANN for generating sequential well-logging data with high accuracy. Consequently, this deep-learning model is applicable to generating synthetic Vp logs at a less consolidated unconventional GH reservoir. We anticipate that the proposed deep-learning approach can be extended to restore or predict well-logging data at missing or unsampled intervals for both conventional and unconventional reservoirs.

Author Contributions: Conceptualization, M.J.; methodology, M.J. and B.M.; software, M.J. and B.M.; validation, M.J., S.K. (Seoyoon Kwon) and B.M.; sensitivity analysis, M.J. and M.K.; investigation, M.J.; writing—original draft, M.J.; writing—review and editing, S.K. (Seoyoon Kwon), M.K. and B.M.; formal Analysis, S.K. (Sungil Kim); supervision, B.M.; project administration, B.M.; funding acquisition, B.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Korea Institute of Geoscience and Mineral Resources (KIGAM) project, "Machine learning-based analysis on gas hydrate (No. GP2021-010)", and the Korea Electric Power Corporation (No. R21XO01-1).

Data Availability Statement: Machine learning data supporting this study's findings are available from the corresponding author upon reasonable request. UBGH field data can only be made available to researchers subject to a non-disclosure agreement due to confidentiality agreements.

Acknowledgments: We are grateful for technical support and software donations from Schlumberger.

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

Appendix A. Training, Validation, and Test Results of the LSTM Model

Tables A1–A5 provide the performance evaluation results visualized in Figure 6.

Table A1. Performance for generating synthetic Vp using a single input variable.

Input	Performance		Number of Neurons in the Hidden Layer							
			1	2	4	8	16	32	64	128
GR	Train	R^2	0.521	0.753	0.771	0.774	0.775	0.774	0.774	0.775
		RMSE	0.487	0.343	0.316	0.312	0.309	0.312	0.309	0.310
	Valid	R^2	0.168	0.409	0.434	0.508	0.518	0.545	0.503	0.481
		RMSE	1.410	1.248	1.193	1.117	1.099	1.078	1.093	1.134
	Test	R^2	0.369	0.539	0.553	0.556	0.556	0.555	0.550	0.554
		RMSE	0.614	0.512	0.499	0.500	0.503	0.501	0.516	0.504

Table A1. *Cont.*

Input	Performance		1	2	4	8	16	32	64	128
			Number of Neurons in the Hidden Layer							
RT	Train	R^2	0.567	0.760	0.779	0.779	0.786	0.790	0.790	0.793
		RMSE	0.447	0.334	0.310	0.307	0.303	0.299	0.300	0.296
	Valid	R^2	0.188	0.267	0.370	0.354	0.485	0.463	0.524	0.480
		RMSE	1.414	1.324	1.251	1.252	1.150	1.172	1.093	1.141
	Test	R^2	0.268	0.535	0.541	0.553	0.554	0.551	0.550	0.548
		RMSE	0.644	0.513	0.506	0.501	0.500	0.506	0.513	0.511
RHOB	Train	R^2	0.424	0.721	0.772	0.778	0.780	0.782	0.779	0.782
		RMSE	0.511	0.363	0.317	0.311	0.306	0.305	0.306	0.304
	Valid	R^2	0.112	0.462	0.538	0.593	0.588	0.528	0.450	0.429
		RMSE	1.445	1.204	1.105	1.045	1.036	1.106	1.156	1.172
	Test	R^2	0.303	0.513	0.557	0.569	0.573	0.570	0.570	0.571
		RMSE	0.632	0.522	0.497	0.491	0.490	0.493	0.499	0.494
TNPH	Train	R^2	0.433	0.740	0.767	0.775	0.775	0.776	0.777	0.780
		RMSE	0.518	0.359	0.318	0.310	0.309	0.308	0.307	0.314
	Valid	R^2	0.266	0.378	0.512	0.491	0.541	0.561	0.562	0.562
		RMSE	1.362	1.267	1.117	1.132	1.077	1.051	1.055	1.087
	Test	R^2	0.369	0.530	0.557	0.562	0.562	0.559	0.560	0.557
		RMSE	0.624	0.516	0.498	0.497	0.498	0.500	0.501	0.497
PEF	Train	R^2	0.530	0.766	0.772	0.773	0.774	0.774	0.773	0.774
		RMSE	0.465	0.337	0.317	0.311	0.309	0.310	0.310	0.310
	Valid	R^2	0.421	0.439	0.518	0.489	0.544	0.591	0.566	0.445
		RMSE	1.274	1.213	1.143	1.126	1.065	1.016	1.043	1.167
	Test	R^2	0.391	0.551	0.558	0.556	0.555	0.555	0.554	0.555
		RMSE	0.589	0.502	0.496	0.502	0.505	0.505	0.505	0.508

Table A2. Performance for generating synthetic Vp using two input variables.

Input	Performance		1	2	4	8	16	32	64	128
			Number of Neurons in the Hidden Layer							
GR RT	Train	R^2	0.827	0.827	0.831	0.836	0.835	0.836	0.844	0.854
		RMSE	0.311	0.293	0.274	0.264	0.264	0.263	0.257	0.249
	Valid	R^2	0.589	0.594	0.602	0.579	0.565	0.551	0.525	0.482
		RMSE	1.126	1.095	1.060	1.055	1.072	1.071	1.098	1.142
	Test	R^2	0.644	0.643	0.640	0.637	0.636	0.639	0.653	0.671
		RMSE	0.460	0.453	0.448	0.451	0.451	0.451	0.443	0.431
GR RHOB	Train	R^2	0.822	0.824	0.833	0.829	0.834	0.837	0.839	0.851
		RMSE	0.314	0.283	0.270	0.271	0.265	0.263	0.261	0.252
	Valid	R^2	0.613	0.626	0.644	0.607	0.621	0.600	0.582	0.558
		RMSE	1.097	1.030	0.992	1.015	0.988	1.013	1.035	1.065
	Test	R^2	0.645	0.646	0.662	0.650	0.658	0.661	0.667	0.683
		RMSE	0.458	0.445	0.434	0.442	0.439	0.439	0.438	0.422
GR TNPH	Train	R^2	0.702	0.827	0.823	0.832	0.837	0.836	0.839	0.843
		RMSE	0.425	0.298	0.281	0.268	0.263	0.266	0.262	0.258
	Valid	R^2	0.259	0.607	0.585	0.633	0.618	0.631	0.604	0.523
		RMSE	1.365	1.080	1.062	0.981	0.996	0.981	0.995	1.080
	Test	R^2	0.548	0.648	0.639	0.655	0.658	0.657	0.661	0.671
		RMSE	0.546	0.449	0.448	0.439	0.439	0.441	0.446	0.439

Table A2. *Cont.*

Input	Performance		Number of Neurons in the Hidden Layer							
			1	**2**	**4**	**8**	**16**	**32**	**64**	**128**
GR PEF	Train	R^2	0.822	0.823	0.824	0.827	0.833	0.830	0.828	0.844
		RMSE	0.316	0.285	0.277	0.273	0.265	0.268	0.270	0.258
	Valid	R^2	0.601	0.585	0.617	0.607	0.599	0.563	0.595	0.499
		RMSE	1.116	1.076	1.025	1.019	1.015	1.050	1.018	1.100
	Test	R^2	0.643	0.644	0.647	0.647	0.656	0.650	0.648	0.679
		RMSE	0.459	0.447	0.444	0.445	0.441	0.447	0.448	0.429
RT RHOB	Train	R^2	0.824	0.829	0.833	0.830	0.836	0.857	0.857	0.854
		RMSE	0.311	0.306	0.274	0.270	0.264	0.246	0.246	0.250
	Valid	R^2	0.609	0.572	0.573	0.592	0.585	0.625	0.600	0.614
		RMSE	1.122	1.123	1.087	1.066	1.051	1.021	1.049	1.040
	Test	R^2	0.629	0.652	0.651	0.638	0.654	0.674	0.673	0.678
		RMSE	0.472	0.451	0.442	0.450	0.441	0.429	0.430	0.431
RT TNPH	Train	R^2	0.825	0.818	0.829	0.840	0.847	0.845	0.846	0.836
		RMSE	0.312	0.295	0.274	0.261	0.256	0.257	0.257	0.268
	Valid	R^2	0.612	0.475	0.514	0.570	0.628	0.615	0.628	0.613
		RMSE	1.103	1.191	1.138	1.078	1.018	1.018	0.987	1.012
	Test	R^2	0.635	0.642	0.640	0.656	0.659	0.670	0.665	0.675
		RMSE	0.465	0.451	0.448	0.438	0.437	0.432	0.441	0.432
RT PEF	Train	R^2	0.695	0.832	0.828	0.840	0.841	0.868	0.863	0.855
		RMSE	0.429	0.282	0.272	0.262	0.261	0.237	0.240	0.248
	Valid	R^2	0.258	0.605	0.567	0.557	0.593	0.617	0.665	0.571
		RMSE	1.366	1.091	1.082	1.103	1.046	1.025	0.958	1.066
	Test	R^2	0.543	0.644	0.645	0.657	0.659	0.691	0.686	0.680
		RMSE	0.549	0.451	0.445	0.438	0.436	0.417	0.420	0.426
RHOB TNPH	Train	R^2	0.696	0.828	0.835	0.834	0.842	0.842	0.844	0.845
		RMSE	0.426	0.283	0.269	0.268	0.260	0.259	0.258	0.256
	Valid	R^2	0.257	0.562	0.586	0.597	0.629	0.565	0.587	0.660
		RMSE	1.365	1.095	1.055	1.036	0.997	1.061	1.021	0.937
	Test	R^2	0.543	0.662	0.671	0.671	0.678	0.672	0.674	0.679
		RMSE	0.546	0.435	0.429	0.429	0.426	0.430	0.437	0.430
RHOB PEF	Train	R^2	0.698	0.828	0.832	0.851	0.844	0.840	0.845	0.847
		RMSE	0.429	0.283	0.270	0.254	0.258	0.261	0.258	0.255
	Valid	R^2	0.256	0.573	0.601	0.638	0.637	0.601	0.658	0.565
		RMSE	1.365	1.085	1.035	1.004	0.995	1.033	0.962	1.068
	Test	R^2	0.549	0.670	0.671	0.692	0.687	0.678	0.689	0.691
		RMSE	0.546	0.430	0.429	0.414	0.419	0.426	0.421	0.418
TNPH PEF	Train	R^2	0.823	0.828	0.826	0.829	0.834	0.833	0.842	0.843
		RMSE	0.311	0.282	0.276	0.270	0.266	0.266	0.259	0.262
	Valid	R^2	0.612	0.622	0.598	0.596	0.590	0.633	0.610	0.612
		RMSE	1.086	1.047	1.052	1.041	1.044	0.980	1.014	1.027
	Test	R^2	0.644	0.650	0.648	0.649	0.656	0.657	0.664	0.664
		RMSE	0.458	0.444	0.443	0.442	0.438	0.441	0.435	0.435

Table A3. Performance for generating synthetic Vp using three input variables.

Input	Performance		Number of Neurons in the Hidden Layer							
			1	2	4	8	16	32	64	128
GR RT RHOB	Train	R^2	0.679	0.805	0.837	0.849	0.864	0.866	0.871	0.872
		RMSE	0.395	0.310	0.272	0.257	0.241	0.240	0.234	0.233
	Valid	R^2	0.291	0.566	0.431	0.524	0.506	0.568	0.572	0.556
		RMSE	1.356	1.164	1.234	1.158	1.148	1.106	1.094	1.100
	Test	R^2	0.481	0.635	0.668	0.684	0.691	0.710	0.709	0.708
		RMSE	0.571	0.466	0.433	0.423	0.416	0.402	0.403	0.404
GR RT TNPH	Train	R^2	0.457	0.687	0.831	0.849	0.858	0.859	0.866	0.867
		RMSE	0.503	0.386	0.277	0.254	0.245	0.245	0.238	0.239
	Valid	R^2	0.277	0.397	0.581	0.640	0.547	0.579	0.599	0.570
		RMSE	1.407	1.302	1.115	1.021	1.108	1.086	1.060	1.099
	Test	R^2	0.233	0.512	0.635	0.665	0.681	0.687	0.692	0.694
		RMSE	0.673	0.551	0.455	0.432	0.422	0.418	0.415	0.414
GR RT PEF	Train	R^2	0.502	0.822	0.838	0.846	0.860	0.866	0.867	0.865
		RMSE	0.481	0.290	0.267	0.260	0.245	0.239	0.240	0.240
	Valid	R^2	0.123	0.610	0.624	0.608	0.626	0.568	0.580	0.574
		RMSE	1.427	1.074	1.052	1.062	1.035	1.088	1.078	1.065
	Test	R^2	0.233	0.636	0.655	0.665	0.674	0.674	0.691	0.689
		RMSE	0.659	0.459	0.440	0.433	0.426	0.427	0.415	0.418
GR RHOB TNPH	Train	R^2	0.492	0.820	0.837	0.850	0.859	0.860	0.859	0.860
		RMSE	0.497	0.308	0.272	0.255	0.244	0.244	0.247	0.251
	Valid	R^2	0.158	0.399	0.557	0.581	0.604	0.661	0.633	0.555
		RMSE	1.407	1.240	1.109	1.080	1.042	0.978	1.012	1.095
	Test	R^2	0.436	0.636	0.683	0.698	0.701	0.703	0.709	0.702
		RMSE	0.603	0.461	0.422	0.410	0.410	0.409	0.403	0.409
GR RHOB PEF	Train	R^2	0.553	0.826	0.821	0.845	0.852	0.859	0.863	0.863
		RMSE	0.452	0.284	0.284	0.258	0.252	0.245	0.242	0.246
	Valid	R^2	0.229	0.593	0.416	0.602	0.529	0.565	0.570	0.490
		RMSE	1.386	1.069	1.226	1.054	1.123	1.078	1.067	1.135
	Test	R^2	0.451	0.673	0.671	0.687	0.696	0.709	0.711	0.713
		RMSE	0.570	0.431	0.431	0.419	0.412	0.404	0.402	0.402
GR TNPH PEF	Train	R^2	0.605	0.802	0.837	0.841	0.857	0.851	0.861	0.861
		RMSE	0.434	0.325	0.272	0.261	0.246	0.254	0.243	0.243
	Valid	R^2	0.274	0.399	0.545	0.629	0.587	0.604	0.578	0.587
		RMSE	1.376	1.276	1.112	1.032	1.051	1.037	1.049	1.035
	Test	R^2	0.502	0.640	0.658	0.671	0.685	0.680	0.691	0.691
		RMSE	0.558	0.472	0.437	0.428	0.421	0.424	0.417	0.420
RT RHOB TNPH	Train	R^2	0.499	0.538	0.817	0.859	0.864	0.872	0.866	0.868
		RMSE	0.478	0.456	0.284	0.246	0.240	0.239	0.238	0.236
	Valid	R^2	0.412	0.382	0.433	0.574	0.568	0.600	0.595	0.540
		RMSE	1.286	1.299	1.208	1.078	1.086	1.068	1.044	1.099
	Test	R^2	0.298	0.286	0.623	0.681	0.688	0.695	0.691	0.686
		RMSE	0.639	0.634	0.461	0.422	0.419	0.412	0.420	0.423
RT RHOB PEF	Train	R^2	0.456	0.653	0.840	0.854	0.861	0.865	0.873	0.873
		RMSE	0.497	0.395	0.267	0.253	0.243	0.239	0.234	0.233
	Valid	R^2	0.232	0.338	0.551	0.499	0.556	0.559	0.607	0.527
		RMSE	1.428	1.302	1.131	1.166	1.104	1.084	1.048	1.120
	Test	R^2	0.267	0.403	0.655	0.683	0.677	0.697	0.703	0.693
		RMSE	0.654	0.583	0.439	0.421	0.425	0.414	0.409	0.419

Table A3. *Cont.*

Input	Performance		Number of Neurons in the Hidden Layer							
			1	2	4	8	16	32	64	128
RT RNPH PEF	Train	R^2	0.541	0.805	0.812	0.849	0.863	0.868	0.865	0.869
		RMSE	0.458	0.308	0.294	0.255	0.242	0.237	0.239	0.236
	Valid	R^2	0.300	0.551	0.490	0.536	0.595	0.549	0.596	0.579
		RMSE	1.351	1.152	1.201	1.116	1.065	1.094	1.047	1.067
	Test	R^2	0.272	0.629	0.648	0.653	0.685	0.696	0.693	0.692
		RMSE	0.642	0.467	0.450	0.440	0.419	0.414	0.415	0.416
RHOB TNPH PEF	Train	R^2	0.662	0.787	0.836	0.857	0.868	0.863	0.859	0.867
		RMSE	0.409	0.322	0.268	0.255	0.238	0.241	0.244	0.238
	Valid	R^2	0.318	0.539	0.573	0.642	0.643	0.573	0.645	0.587
		RMSE	1.360	1.176	1.068	1.012	0.990	1.056	0.991	1.052
	Test	R^2	0.518	0.602	0.671	0.691	0.705	0.700	0.695	0.701
		RMSE	0.535	0.484	0.428	0.417	0.406	0.412	0.416	0.410

Table A4. Performance for generating synthetic Vp using four input variables.

Input	Performance		Number of Neurons in the Hidden Layer							
			1	2	4	8	16	32	64	128
GR RT RHOB TNPH	Train	R^2	0.657	0.893	0.887	0.908	0.921	0.923	0.929	0.925
		RMSE	0.420	0.260	0.225	0.200	0.183	0.180	0.173	0.179
	Valid	R^2	0.354	0.517	0.593	0.715	0.759	0.703	0.703	0.660
		RMSE	1.319	1.181	1.086	0.956	0.857	0.939	0.941	0.985
	Test	R^2	0.475	0.787	0.794	0.810	0.835	0.832	0.842	0.846
		RMSE	0.590	0.370	0.343	0.326	0.303	0.306	0.298	0.294
GR RT RHOB PEF	Train	R^2	0.877	0.895	0.900	0.917	0.921	0.927	0.924	0.916
		RMSE	0.281	0.234	0.216	0.190	0.183	0.178	0.179	0.192
	Valid	R^2	0.650	0.718	0.708	0.709	0.740	0.732	0.696	0.738
		RMSE	1.096	1.015	0.976	0.942	0.909	0.929	0.959	0.894
	Test	R^2	0.771	0.783	0.795	0.817	0.828	0.836	0.838	0.844
		RMSE	0.397	0.363	0.344	0.319	0.310	0.303	0.301	0.296
GR RT TNPH PEF	Train	R^2	0.896	0.892	0.908	0.911	0.925	0.925	0.927	0.931
		RMSE	0.266	0.236	0.205	0.196	0.180	0.180	0.177	0.171
	Valid	R^2	0.691	0.734	0.649	0.692	0.754	0.731	0.689	0.689
		RMSE	1.049	0.932	1.021	0.956	0.879	0.919	0.948	0.931
	Test	R^2	0.780	0.785	0.802	0.815	0.834	0.844	0.846	0.847
		RMSE	0.389	0.360	0.337	0.323	0.304	0.296	0.294	0.294
GR RHOB TNPH PEF	Train	R^2	0.894	0.910	0.911	0.907	0.929	0.930	0.930	0.926
		RMSE	0.260	0.215	0.206	0.201	0.174	0.174	0.171	0.177
	Valid	R^2	0.654	0.788	0.754	0.691	0.773	0.711	0.711	0.747
		RMSE	1.076	0.856	0.897	0.943	0.797	0.871	0.884	0.830
	Test	R^2	0.784	0.820	0.828	0.810	0.843	0.845	0.848	0.840
		RMSE	0.374	0.325	0.315	0.326	0.297	0.299	0.294	0.291
RT RHOB TNPH PEF	Train	R^2	0.896	0.896	0.901	0.905	0.911	0.918	0.923	0.923
		RMSE	0.262	0.231	0.213	0.201	0.195	0.186	0.181	0.181
	Valid	R^2	0.648	0.692	0.712	0.643	0.729	0.728	0.701	0.712
		RMSE	1.091	0.988	0.940	0.985	0.886	0.900	0.918	0.894
	Test	R^2	0.779	0.779	0.786	0.790	0.808	0.826	0.833	0.832
		RMSE	0.383	0.362	0.348	0.342	0.328	0.312	0.306	0.310

Table A5. Performance for generating synthetic Vp using five input variables.

Input	Performance		Number of Neurons in the Hidden Layer							
			1	2	4	8	16	32	64	128
GR RT RHOB TNPH PEF	Train	R^2	0.894	0.898	0.910	0.918	0.923	0.914	0.924	0.925
		RMSE	0.263	0.242	0.205	0.190	0.181	0.192	0.180	0.178
	Valid	R^2	0.632	0.692	0.744	0.714	0.766	0.724	0.645	0.673
		RMSE	1.109	1.013	0.937	0.933	0.848	0.921	0.999	0.960
	Test	R^2	0.783	0.788	0.806	0.816	0.828	0.810	0.833	0.837
		RMSE	0.377	0.361	0.332	0.320	0.310	0.325	0.306	0.303

References

1. Chand, S.; Minshull, T.A. Seismic Constraints on the Effects of Gas Hydrate on Sediment Physical Properties and Fluid Flow: A Review. *Geofluids* **2003**, *3*, 275–289. [CrossRef]
2. Liang, H.; Guan, D.; Shi, K.; Yang, L.; Zhang, L.; Zhao, J.; Song, Y. Characterizing Mass-Transfer Mechanism during Gas Hydrate Formation from Water Droplets. *Chem. Eng. J.* **2022**, *428*, 132626. [CrossRef]
3. Pang, X.-Q.; Chen, Z.-H.; Jia, C.-Z.; Wang, E.-Z.; Shi, H.-S.; Wu, Z.-Y.; Hu, T.; Liu, K.-Y.; Zhao, Z.-F.; Pang, B.; et al. Evaluation and Re-Understanding of the Global Natural Gas Hydrate Resources. *Pet. Sci.* **2021**, *18*, 323–338. [CrossRef]
4. Wang, T.; Hu, T.; Pang, X.-Q.; Zhang, X.-W.; Liu, X.-H.; Xu, Z.; Wang, E.-Z.; Wu, Z.-Y. Distribution and Resource Evaluation of Natural Gas Hydrate in South China Sea by Combing Phase Equilibrium Mechanism and Volumetric Method. *Pet. Sci.* **2022**, *19*, 26–36. [CrossRef]
5. Yang, J.; Lu, M.; Yao, Z.; Wang, M.; Lu, S.; Qi, N.; Xia, Y. A Geophysical Review of the Seabed Methane Seepage Features and Their Relationship with Gas Hydrate Systems. *Geofluids* **2021**, *2021*, 9953026. [CrossRef]
6. Yoon, H.C.; Yoon, S.; Lee, J.Y.; Kim, J. Multiple Porosity Model of a Heterogeneous Layered Gas Hydrate Deposit in Ulleung Basin, East Sea, Korea: A Study on Depressurization Strategies, Reservoir Geomechanical Response, and Wellbore Stability. *J. Nat. Gas Sci. Eng.* **2021**, *96*, 104321. [CrossRef]
7. Gornitz, V.; Fung, I. Potential Distribution of Methane Hydrates in the World's Oceans. *Global Biogeochem. Cycles* **1994**, *8*, 335–347. [CrossRef]
8. Boswell, R.; Collett, T.S. Current Perspectives on Gas Hydrate Resources. *Energy Environ. Sci.* **2011**, *4*, 1206–1215. [CrossRef]
9. Meng, M.; Liang, J.; Lu, J.; Zhang, W.; Kuang, Z.; Fang, Y.; He, Y.; Deng, W.; Huang, W. Quaternary Deep-Water Sedimentary Characteristics and Their Relationship with the Gas Hydrate Accumulations in the Qiongdongnan Basin, Northwest South China Sea. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **2021**, *177*, 103628. [CrossRef]
10. Zhu, X.; Liu, X. Morphology Identification of Gas Hydrate from Pointwise Lipschitz Regularity for P- and S-Wave Velocity. *Energy Rep.* **2021**, *7*, 8062–8074. [CrossRef]
11. Saggaf, M.M.; Nebrija, E.L. Estimation of Missing Logs by Regularized Neural Networks. *Am. Assoc. Pet. Geol. Bull.* **2003**, *87*, 1377–1389. [CrossRef]
12. Bader, S.; Wu, X.; Fomel, S. Missing Log Data Interpolation and Semiautomatic Seismic Well Ties Using Data Matching Techniques. *Interpretation* **2019**, *7*, T347–T361. [CrossRef]
13. Onalo, D.; Adedigba, S.; Khan, F.; James, L.A.; Butt, S. Data Driven Model for Sonic Well Log Prediction. *J. Pet. Sci. Eng.* **2018**, *170*, 1022–1037. [CrossRef]
14. Wang, P.; Peng, S. On a New Method of Estimating Shear Wave Velocity from Conventional Well Logs. *J. Pet. Sci. Eng.* **2019**, *180*, 105–123. [CrossRef]
15. Dalvand, M.; Falahat, R. A New Rock Physics Model to Estimate Shear Velocity Log. *J. Pet. Sci. Eng.* **2021**, *196*, 107697. [CrossRef]
16. Gharbi, R.B.C.; Mansoori, G.A. An Introduction to Artificial Intelligence Applications in Petroleum Exploration and Production. *J. Pet. Sci. Eng.* **2005**, *49*, 93–96. [CrossRef]
17. ZHANG, D.; CHEN, Y.; MENG, J. Synthetic Well Logs Generation via Recurrent Neural Networks. *Pet. Explor. Dev.* **2018**, *45*, 629–639. [CrossRef]
18. Tang, L.; Li, J.; Lu, W.; Lian, P.; Wang, H.; Jiang, H.; Wang, F.; Jia, H. Well Control Optimization of Waterflooding Oilfield Based on Deep Neural Network. *Geofluids* **2021**, *2021*, 8873782. [CrossRef]
19. Ji, M.; Kwon, S.; Park, G.; Min, B.; Nguyen, X.H. Prediction of Water Saturation from Well Log Data Using Deep Learning Algorithms. *J. Korean Soc. Miner. Energy Resour. Eng.* **2021**, *58*, 215–226. [CrossRef]
20. Shewalkar, A.; Nyavanandi, D.; Ludwig, S.A. Performance Evaluation of Deep Neural Networks Applied to Speech Recognition: RNN, LSTM and GRU. *J. Artif. Intell. Soft Comput. Res.* **2019**, *9*, 235–245. [CrossRef]
21. Kim, S.; Kim, K.H.; Min, B.; Lim, J.; Lee, K. Generation of Synthetic Density Log Data Using Deep Learning Algorithm at the Golden Field in Alberta, Canada. *Geofluids* **2020**, *2020*, 5387183. [CrossRef]
22. Pham, N.; Wu, X.; Zabihi Naeini, E. Missing Well Log Prediction Using Convolutional Long Short-Term Memory Network. *GEOPHYSICS* **2020**, *85*, WA159–WA171. [CrossRef]

23. Zhang, Y.; Zhong, H.-R.; Wu, Z.-Y.; Zhou, H.; Ma, Q.-Y. Improvement of Petrophysical Workflow for Shear Wave Velocity Prediction Based on Machine Learning Methods for Complex Carbonate Reservoirs. *J. Pet. Sci. Eng.* **2020**, *192*, 107234. [CrossRef]
24. Lee, M.W.; Waite, W.F. Estimating Pore-Space Gas Hydrate Saturations from Well Log Acoustic Data. *Geochem. Geophys. Geosystems* **2008**, *9*, 7. [CrossRef]
25. Haines, S.S.; Collett, T.; Boswell, R.; Lim, T.K.; Okinaka, N.; Suzuki, K.; Fujimoto, A. Gas Hydrate Saturation Estimation from Acoustic Log Data in the 2018 Alaska North Slope Hydrate-01 Stratigraphic Test Well. In Proceedings of the 10th International Conference on Gas Hydrates (ICGH10), Singapore, 21–26 June 2020.
26. Helgerud, M.B.; Dvorkin, J.; Nur, A.; Sakai, A.; Collett, T. Elastic-Wave Velocity in Marine Sediments with Gas Hydrates: Effective Medium Modeling. *Geophys. Res. Lett.* **1999**, *26*, 2021–2024. [CrossRef]
27. Lee, M.W.; Collett, T.S. In-Situ Gas Hydrate Hydrate Saturation Estimated from Various Well Logs at the Mount Elbert Gas Hydrate Stratigraphic Test Well, Alaska North Slope. *Mar. Pet. Geol.* **2011**, *28*, 439–449. [CrossRef]
28. You, J.; Cao, J.; Wang, X.; Liu, W. Shear Wave Velocity Prediction Based on LSTM and Its Application for Morphology Identification and Saturation Inversion of Gas Hydrate. *J. Pet. Sci. Eng.* **2021**, *205*, 109027. [CrossRef]
29. Chen, Y.; Zhang, D. Physics-Constrained Deep Learning of Geomechanical Logs. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5932–5943. [CrossRef]
30. Naim, F.; Cook, A. Estimating P-Wave Velocity and Bulk Density in Hydrate Systems Using Machine Learning. In Proceedings of the AGU Fall Meeting 2021, New Orleans, LA, USA, 13–17 December 2021; Volume 2021, p. OS22A-07.
31. Karimpouli, S.; Tahmasebi, P. Physics Informed Machine Learning: Seismic Wave Equation. *Geosci. Front.* **2020**, *11*, 1993–2001. [CrossRef]
32. Huh, D.-G.; Lee, J.Y. Overview of Gas Hydrates R&D. *J. Korean Soc. Miner. Energy Resour. Eng.* **2017**, *54*, 201–214. [CrossRef]
33. Kim, G.Y.; Yi, B.Y.; Yoo, D.G.; Ryu, B.J.; Riedel, M. Evidence of Gas Hydrate from Downhole Logging Data in the Ulleung Basin, East Sea. *Mar. Pet. Geol.* **2011**, *28*, 1979–1985. [CrossRef]
34. Lee, M.W.; Collett, T.S. Gas Hydrate and Free Gas Saturations Estimated from Velocity Logs on Hydrate Ridge, Offshore Oregon, USA. In *Proceedings of the Ocean Drilling Program, 199 Scientific Results*; Ocean Drilling Program: College Station, TX, USA, 2006.
35. Moridis, G.J.; Kim, J.; Reagan, M.T.; Kim, S.-J. Feasibility of Gas Production from a Gas Hydrate Accumulation at the UBGH2-6 Site of the Ulleung Basin in the Korean East Sea. *J. Pet. Sci. Eng.* **2013**, *108*, 180–210. [CrossRef]
36. Lee, T.; Lee, J.Y.; Ahn, T.; Son, H.A. Numerical Simulation of Gas Hydrate Production Using the Cyclic Depressurization Method in the Ulleung Basin of the Korea East Sea. *Appl. Sci.* **2021**, *11*, 9748. [CrossRef]
37. Park, S.Y.; Son, B.-K.; Choi, J.; Jin, H.; Lee, K. Application of Machine Learning to Quantification of Mineral Composition on Gas Hydrate-Bearing Sediments, Ulleung Basin, Korea. *J. Pet. Sci. Eng.* **2022**, *209*, 109840. [CrossRef]
38. Ryu, B.-J.; Collett, T.S.; Riedel, M.; Kim, G.Y.; Chun, J.-H.; Bahk, J.-J.; Lee, J.Y.; Kim, J.-H.; Yoo, D.-G. Scientific Results of the Second Gas Hydrate Drilling Expedition in the Ulleung Basin (UBGH2). *Mar. Pet. Geol.* **2013**, *47*, 1–20. [CrossRef]
39. Lee, G.H.; Kim, H.J.; Suh, M.C.; Hong, J.K. Crustal Structure, Volcanism, and Opening Mode of the Ulleung Basin, East Sea (Sea of Japan). *Tectonophysics* **1999**, *308*, 503–525. [CrossRef]
40. Horozal, S.; Bahk, J.-J.; Urgeles, R.; Kim, G.Y.; Cukur, D.; Kim, S.-P.; Lee, G.H.; Lee, S.H.; Ryu, B.-J.; Kim, J.-H. Mapping Gas Hydrate and Fluid Flow Indicators and Modeling Gas Hydrate Stability Zone (GHSZ) in the Ulleung Basin, East (Japan) Sea: Potential Linkage between the Occurrence of Mass Failures and Gas Hydrate Dissociation. *Mar. Pet. Geol.* **2017**, *80*, 171–191. [CrossRef]
41. Horozal, S.; Kim, G.Y.; Bahk, J.J.; Wilkens, R.H.; Yoo, D.G.; Ryu, B.J.; Kim, S.P. Core and Sediment Physical Property Correlation of the Second Ulleung Basin Gas Hydrate Drilling Expedition (UBGH2) Results in the East Sea (Japan Sea). *Mar. Pet. Geol.* **2015**, *59*, 535–562. [CrossRef]
42. Honarkhah, M.; Caers, J. Stochastic Simulation of Patterns Using Distance-Based Pattern Modeling. *Math. Geosci.* **2010**, *42*, 487–517. [CrossRef]
43. Shin, H.-J.; Lim, J.-S.; Kim, S.-J. Estimation of Porosity and Saturation in Gas Hydrate Bearing Sediments Using Well Logs and Core Analysis Data of the 2nd Wells in Ulleung Basin, East Sea. *Korea, J. Korean Soc. Miner. Energy Resour. Eng.* **2012**, *49*, 175–185.
44. Bahk, J.-J.; Kim, G.-Y.; Chun, J.-H.; Kim, J.-H.; Lee, J.Y.; Ryu, B.-J.; Lee, J.-H.; Son, B.-K.; Collett, T.S. Characterization of Gas Hydrate Reservoirs by Integration of Core and Log Data in the Ulleung Basin, East Sea. *Mar. Pet. Geol.* **2013**, *47*, 30–42. [CrossRef]
45. Kim, A.-R.; Kim, H.-S.; Cho, G.-C.; Lee, J.Y. Estimation of Model Parameters and Properties for Numerical Simulation on Geomechanical Stability of Gas Hydrate Production in the Ulleung Basin, East Sea, Korea. *Quat. Int.* **2017**, *459*, 55–68. [CrossRef]
46. Kim, J.-T.; Kang, C.-W.; Kim, A.-R.; Lee, J.Y.; Cho, G.-C. Effect of Permeability on Hydrate-Bearing Sediment Productivity and Stability in Ulleung Basin, East Sea, South Korea. *Energies* **2021**, *14*, 1752. [CrossRef]
47. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshad, H. State-of-the-Art in Artificial Neural Network Applications: A Survey. *Heliyon* **2018**, *4*, e00938. [CrossRef] [PubMed]
48. Kwon, S.; Park, G.; Jang, Y.; Cho, J.; Chu, M.; Min, B. Determination of Oil Well Placement Using Convolutional Neural Network Coupled with Robust Optimization under Geological Uncertainty. *J. Pet. Sci. Eng.* **2021**, *201*, 108118. [CrossRef]
49. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
50. Lee, H.W.; Kim, M.; Son, H.W.; Min, B.; Choi, J.H. Machine-Learning-Based Water Quality Management of River with Serial Impoundments in the Republic of Korea. *J. Hydrol. Reg. Stud.* **2022**, *41*, 101069. [CrossRef]

51. Asoodeh, M.; Bagheripour, P. Prediction of Compressional, Shear, and Stoneley Wave Velocities from Conventional Well Log Data Using a Committee Machine with Intelligent Systems. *Rock Mech. Rock Eng.* **2012**, *45*, 45–63. [CrossRef]
52. Zoveidavianpoor, M.; Samsuri, A.; Shadizadeh, S.R. Adaptive Neuro Fuzzy Inference System for Compressional Wave Velocity Prediction in a Carbonate Reservoir. *J. Appl. Geophys.* **2013**, *89*, 96–107. [CrossRef]
53. Sun, X.; Liu, Y.; Li, J.; Zhu, J.; Chen, H.; Liu, X. Feature Evaluation and Selection with Cooperative Game Theory. *Pattern Recognit.* **2012**, *45*, 2992–3002. [CrossRef]
54. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
55. Sheela, K.G.; Deepa, S.N. Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Math. Probl. Eng.* **2013**, *2013*, 425740. [CrossRef]
56. Orabona, F.; Tommasi, T. Training Deep Networks without Learning Rates Through Coin Betting. In *Advances in Neural Information Processing Systems*; Guyon, I., von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
57. Kandel, I.; Castelli, M. The Effect of Batch Size on the Generalizability of the Convolutional Neural Networks on a Histopathology Dataset. *ICT Express* **2020**, *6*, 312–315. [CrossRef]