Science & Technology Development Journal – Engineering and Technology, 4(SI3):SI42-SI48

Open Access Full Text Article

Original Research

Testing neural networks assessment based on data-driven using well log data in Cuu Long basin, offshore Vietnam

Nguyen Xuan Huy[1,2,*], Nguyen Thi Thu Trang[1,2], Le Hoang Vu[1,2], Tran Ngoc Bao Duy[2,3], Tran Pham Vu[2,3], Ta Quoc Dung[1,2]

[1]*Faculty of Geology and Petroleum Engineering , Ho Chi Minh University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam*

[2]*Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc District, Ho Chi Minh City,Vietnam*

[3]*Department of Computer Science, Ho Chi Minh University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam*

Correspondence

Nguyen Xuan Huy, Faculty of Geology and Petroleum Engineering , Ho Chi Minh University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Vietnam

Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc District, Ho Chi Minh City,Vietnam

Email: nxhuy@hcmut.edu.vn

History

- Received: 09-9-2019
- Accepted: 11-10-2021
- Published: 03-11-2021

DOI : 10.32508/stdjet.v4iSI3.584

ABSTRACT

This study evaluates the effectiveness of neural network testing on well-log data in the study area. The Artificial Neural Networks (ANNs) and Convolutional neural networks (CNNs) models are developed to predict the missing part of the data or verify the values due to errors in the measurement process. In addition, neural networks are also used to create virtual logs at any location in the reservoir based on log data from existing wells to get a better view of the geological characteristics in the subsurface without any new drilling wells. The dataset used in this study includes qualified logs in twenty wells located in Cuu Long Basin. Data sets for neural networks are designed based on the characteristics of the log data, including the direction of the target well, the angle of the goal well, the position, the depth, and the log values of the nearest wells. Min-max normalization is used to scale the well length before training the dataset. The database is divided into three different sets: training data set, test set validation data set, and test data set. The reliability and accuracy of the methods are expressed through the loss function or the correlation coefficient R^2. The accuracy of these logs was tested for newly drilled wells at the time the system was developed and trained. Log values generated by CNNs have higher correlation coefficients than those of ANNs with R^2 equal to 0.7994, while R^2 of ANNs is only 0.6701. Results showed that predicting using CNNs was better than ANNs. Therefore, the use of CNNs will increase decision-making efficiency by avoiding time-consuming procedures and processes.

Key words: Artificial neural networks, Convolutional neural networks, well log, loss function

INTRODUCTION

Log data in most wells are usually available to provide geological information in the subsurface across the well. With the advancement of artificial intelligence in recent years, many studies have been done to solve the problem of well-log data using neural networks. McCormack[1] introduced artificial neural networks in 1991 for lithological characterization from resistivity log and potential spontaneous log. Rezaee[2] introduced an approach using artificial intelligence to synthesize geophysical log curves in southern Iran. The authors used fuzzy logic and FCNN (Fully Connected Neural Network) to create a log curve for the 3rd well, with input data from 2 other wells[2]. In 2011, Ghavami[3] introduced a method to generate artificial log curves using data from 5 wells, including log gamma, log density, log neutron and log well radius. The data of the first 4 wells are used to train the development and training of FCNN neural networks, and the data of the remaining (5^{th}) wells are used for testing purposes [3].

Long[4] presented an automated process including data processing, data mining techniques and application of artificial neural network to predict and generate density logs. Thirty-eight log types of each of the 12 wells were used to generate log densities for another well called pairwise prediction[4]. Salehi[5] introduced a hidden multi-layer neural network to develop a predictive model of log lines. Traditional log curves such as the porosity log and the sonic log are used to predict the density and resistivity of the deep layer[5].

Inspired by previous publications, this study also contributes to testing the performance of neural networks on well log data. An effective and inexpensive data-driven method is presented to predict a new log set at any location in the field using neural networks, namely ANNs and CNNs. This method is demonstrated on log data set in one oil field in Cuu Long Basin with relatively good quality.

METHODOLOGY

Artificial Neural Networks (ANNs)

An Artificial Neural Network (ANN) is an information processing model that is simulated based on the activity of an organism's neural system, consisting

Cite this article : Huy N X, Trang N T T, Vu L H, Duy T N B, Vu T P, Dung T Q. **Testing neural networks assessment based on data-driven using well log data in Cuu Long basin, offshore Vietnam**. *Sci. Tech. Dev. J. – Engineering and Technology;* 4(SI3):SI42-SI48.

of a large number of neurons connected to information processing. ANN is like the human brain, learned by experience (through training), capable of storing learned experiences (knowledge) and using that knowledge in predicting unknown data (unseen data)[6].

A typical ANN network architecture usually consists of an input layer, one or more hidden layers, and an output layer, as shown in Figure 1[7]. As illustrated in Figure 1, the first layer is the input layer, the middle layer is called the hidden layer, and the last layer is the output layer. Input from the node in the ith layer will receive inputs from the nodes in the i-1th layer and produce output in the i+1th layer[7]. The connection between input and output is established through internal computations in hidden layers. The circles illustrated in Figure 1 are called nodes, also called neurons[7]. The complexity and nonlinearity of the model are increased by increasing the number of hidden layers. During the training of the ANN model, all the weights and bias coefficients are determined by minimizing the error between the prediction output and the training output through the activation function at each node. The activation function acts as the nonlinear component at the output of the neurons in the ANN[8]. There are four activation functions: sigmoid, hyperbolic tangent, linear, and rectified linear unit (ReLU). The activation function is usually a function applied to each element of the input matrix or vector; in other words, the activation function is usually element-wise.

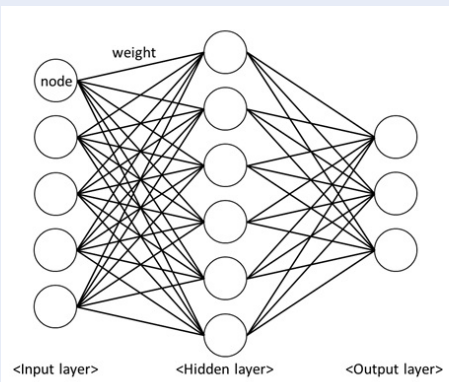

Figure 1: Typical illustration structure of ANN[7].

The mathematical equation of the neural network is written in the following form:

$$Y = \sigma(WX + B) \quad (1)$$

Where Y is the output; σ is the activation function; W is the weight; X is the input, and B is the bias.

Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) are generally deep ANNs formulated from image processing and computer vision. More recently, convolutional networks have been applied directly to text analytics and graph data. The efficacy of convolutional nets (ConvNets or CNNs) in image recognition is one of the main reasons why the world has woken up to the effectiveness of deep learning. Nowadays, many researchers applied this methodology to many fields of computing.

Based on the principle of ANNs, CNNs receive an input (a single vector) and transform it through a series of *hidden layers*. Each hidden layer is made up of a set of neurons, where each neuron is fully connected to all neurons in the previous segment, and where neurons in a single layer function entirely independently and do not share any connections. The last fully-connected layer is called the “output layer” and in classification settings, it represents the class scores[9].

However, the propagation and backpropagation in CNNs have a considerable difference in computing. In summary, there are some following significant steps[10]:

- Convolution: Convolutional filters (kernels) extract features from an input that can be invisible to humans. The filter read successively, from left to right and from top to bottom, all the “pixels” (like a cell in the input matrix) of the kernel action area (as illustrated in Figure 2).
- Pooling: After obtaining features using convolution for describing a large input, downsize the feature map and highlight main features for improving the training performance of neural networks (Figure 3)
- Fully connected layer: Neurons in a fully connected layer have full connections to all activations in the previous segment, as seen in regular Neural Networks. Their activations can hence be computed with a matrix multiplication followed by a bias offset.

The most fundamental advantage of a convolutional neural network is automatic feature extraction for the given task, provided that the input can be represented as a tensor in which local elements are correlated with one another. The functional form of this neural network tends to make no assumptions on the input vector that is supplied to the first layer other than the idea that each element is meaningful to the task.

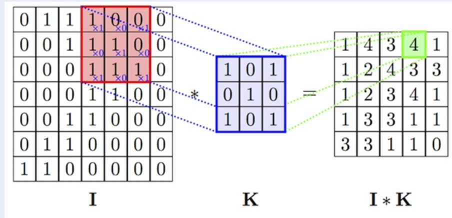

Figure 2: An example of convolution operation in 2D

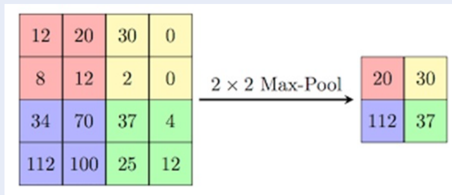

Figure 3: An example of max-pooling in 2D

In general, when using such a network, a machine learning designer will need to think about relevant features for the given task and assign them to an input vector element [11]. On the other hand, a convolutional neural network will automatically extract such features given that you can at least represent the input as a tensor in which elements are locally correlated.

RESULTS AND DISCUSSION

Dataset

Over the twenty's wells located in Cuu Long Basin were chosen for this study. Most of the well contain qualified logs consisting of SP, GR, NPHI, DTCO, DTSM, Res Deep, Res Shallow… and available in the database.

Normalization

Min-max normalization performs linear normalization on the original data. Suppose min A and max A is the smallest and largest values of attribute A. Standardized min-max will map the value v of A to v' value within [new minA, new_maxA] with the following formula:

$$v' = \frac{v - min}{max - min}(newmax - newmin) + newmin \qquad (2)$$

Min-max normalization ensures the relationship between the values of the original data. It will detect errors that "exceed the limit" if input data in the future falls outside the initial value range of A.

Each normalized log data in the range [0, 1] will ensure that all values in the data set are adjusted. Data will be shift and scale through the minimum and range of the wells' data, respectively, representing a 95% confidence interval. Gaussian distribution is assumed for each data set. Normalized values will allow a clear comparison with each corresponding feature in the CNNs.

Each log-in well will have different measurement depths. To train neural networks with a data set, the size of the input data sets and the size of the output data sets must be the same for all training pairs. Therefore, a scale of well length should be converted into an average well length with several specific data points covering the entire area of interest.

Figure 4 shows standardized training wells and testing wells.

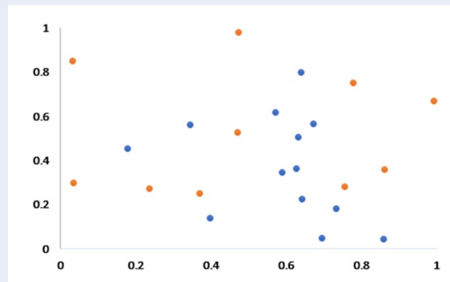

Figure 4: Location of training well (blue dot) and testing well (red dot) in the study area after adopting min-max normalization.

Training, Validation and Testing Data Set

To ensure confidence in the result of prediction from wells in new geological areas, the database is divided into three different sets, called training data set, test set validation data set, and test data set.

The training data set is used to create a model that fits the data. If the model is too complicated in terms of calculation, or little data, it will lead to overtraining. Then, the model will be useful on the training data set, however, when the application enters the new data set, it does not bring the desired results.

The validation data set includes data that the model has never been trained before and is used to evaluate the generalization of the model. The model will not meet the validation data set immediately, but indirectly, many prediction decisions will be made based on the results of the model. When the error of the model on the training set and the validation set is not too large, the model is considered to have achieved certain efficiency and can be regarded as general ability.

A third data set, called a testing data set, is used to independently re-test the model's results.

This study used data from 13 wells for training with 214577 data points, and ten wells were used as validation sets. And only the wells trained are used for neural network testing to avoid overfitting.

Data sets for neural networks are designed based on the characteristics of the log data, including the direction of the target well, the angle of the goal well, the position, the depth, and the log values of the nearest wells. The hidden layers of the network have 100, 50, 3 neurons, respectively.

Each neural network is trained for 1000 epochs, where it represents one pass of the full training set.

Loss function J and coefficient of determination R^2 was used to evaluate the performance of ANN and CNN. In essence, the loss function is a function that allows determining the degree of deviation of the prediction result from the actual price to be predicted. It is a method of measuring the quality of a predictive model on an observational data set. If the prediction model is wrong, the value of the loss function will be larger and vice versa; if it is correct, the loss function value will be lower. The loss function is calculated as follows[7]:

$$J = RMSE = \sqrt{\frac{1}{N_{train}} \sum_{i=1}^{N_{train}} \left(Y_{Nfc} - \widehat{Y}\right)_i^2}$$

Where N_{train} is the number of training data

The coefficient R^2 in the range 0-1 indicates the correlation between the results from the neural network and the reference output[7]. The closer R^2 is to 1, the higher the performance of the network and vice versa. In general, the network is selected when a minimum validation error is obtained to prevent overfitting. Figure 5 illustrates errors in neural networks training of ANNs and CNNs. As illustrated in Figure 5, ANN has a smaller training error than the training error of CNN. However, CNN shows more efficient prediction ability than ANN on validation test, with much smaller validation errors than those of ANN. This shows that the ANN has been overfitted during training, leading to many errors when performing on the validation set.

Figure 6 shows the scatter-plot that compares the test of ANNs (Figure 6a) and corresponding CNNs (Figure 6b). The closer the data points (shown in red on the graph) to the reference line (blue line), the more accurate the performance of the neural network. In Figure 6, R^2 of CNNs (0.7994) is higher than R^2 of the ANNs (0.6701) that proves CNN is more effective than ANN.

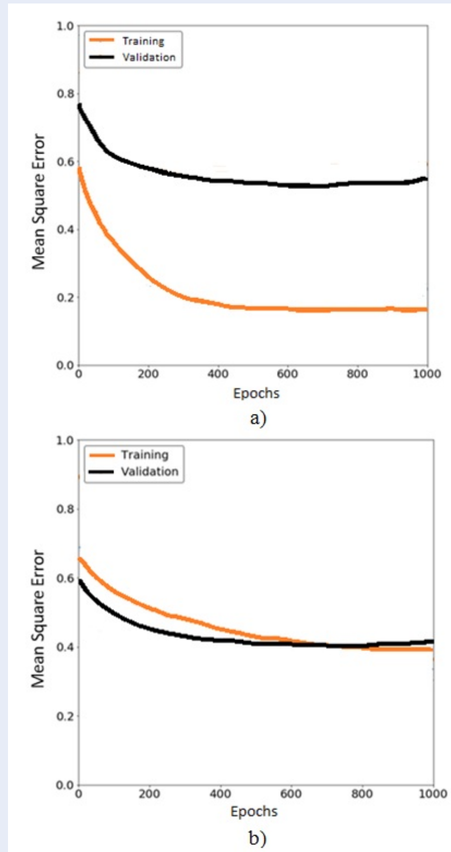

Figure 5: Model losses during training for (a) ANN and (b) CNN

Results and Discussion

The result of trained neural networks has been used to create a log or missing data from the well or due to errors in the measurement of physical devices.

Figure 7 shows the performance of ANNs and CNNs in filling the missing part of the log due to measurement errors. The initial values (actual values) in the depth range from 1825m – 1925m are out of the range of the NPHI log, and neural networks used to validate the NPHI log. Both ANN and CNN show the ability to recover log values in this depth range, but CNN gives more smooth values with a higher correlation coefficient while ANN is suffered from overfitting (as mentioned in Figure 5 and Figure 6).

The most powerful application of this neural network is to create a predicted log at any location in the study area. The results of the neural network model were built for predicting log NPHI for the new wells using ANNs and CNNs shown in Figure 8. The NPHI log in the blue line in Figure 8 shows the actual values obtained from the new drilling well. The real values

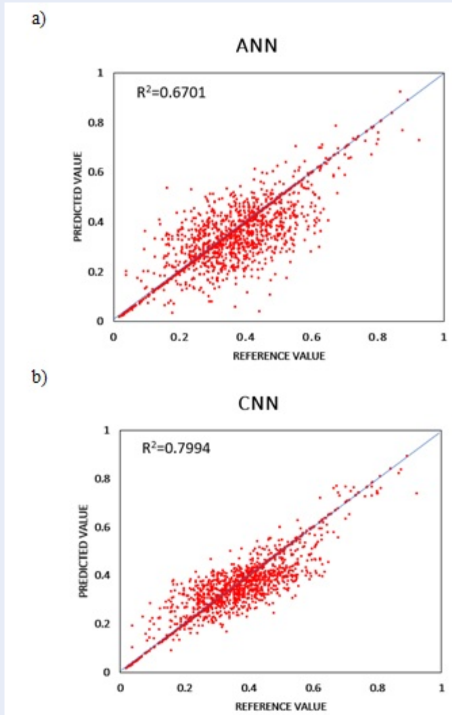

Figure 6: Scatter-plot compares test results of NPHI log for: a) ANN b) CNN.

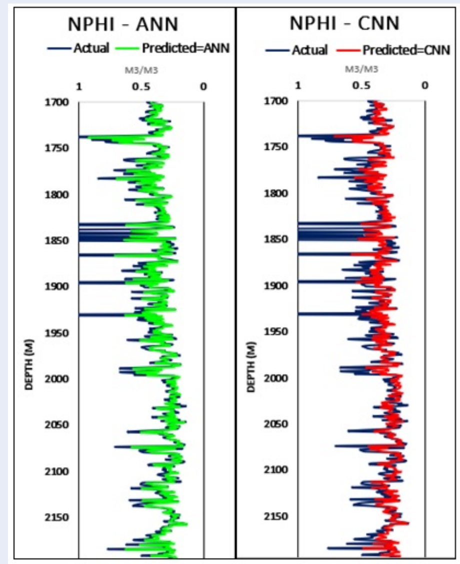

Figure 7: Filling the missing part of the log due to errors while measuring and verifying log values using neural networks.

of this new well are used to test the accuracy of predicted logs by ANN and CNN at the time the system was developed and trained. Similarly, both ANN and CNN can predict log for a new well, but CNN is more efficient and more reliable because CNN does not occur overfitting as ANN does (as mentioned in Figure 5 and Figure 6).

CONCLUSIONS

The performance of ANNs shows overfitting due to an increase in validation errors, despite smaller training errors than those of the CNN. Log values generated by CNNs have higher correlation coefficients than those of ANNs with R^2 equal to 0.7994, while R^2 of ANNs is only 0. 6701. The trial run of CNNs will result in a more accurate evaluation than ANNs, with the reliability is also higher than that of ANNs.

Therefore, the use of CNNs increases the efficiency in restoring missing parts of logs on well log data and helping to predict logs for new wells at any location from available log data sets in existing wells without drilling any new wells.

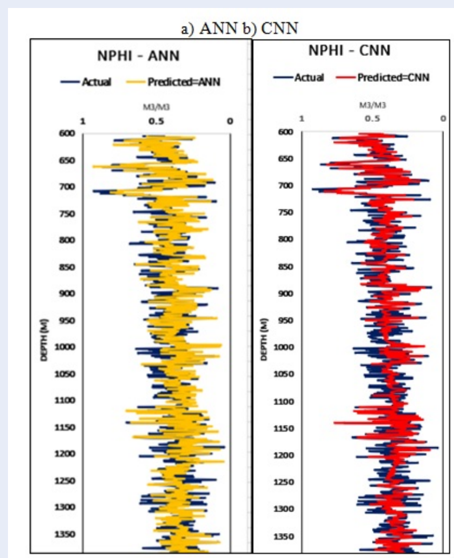

Figure 8: Predict the log NPHI using the ANN (a) and CNN (b) tests for a new well and compare the results with the actual well.

ACKNOWLEDGMENT

This research is funded by Ho Chi Minh City University of Technology–VNU-HCM, under grant number

To-ĐCDK-2020-14. In addition, this research is also funded by the Ministry of Science and Technology of Vietnam, under grant number NĐT.48.KR/18.
We would like to thank Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for the support of time and facilities for this study.

A BBREVIATION

B Bias array
I Number of input array types
J Loss function
N_{train} Number of training data
W Convolutional filter or weight array
X Input array
Y Output array
N_{Nfc} Final output
Activation function
ReLU Rectified Linear Unit
FCNN Fully Connected Neural Network
ANN Artificial Neural Network
ANNs Artificial Neural Networks
CNN Convolutional Neural Network
CNNs Convolutional Neural Networks
SP Spontaneous Potential
GR Gamma Ray
NPHI Neutron Porosity
DTCO Compressional wave delay time
DTSM Shear wave delay time

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTION

Huy Nguyen Xuan: Funding acquisition, Data curation, Supervision, Writing – review and editting
Trang Nguyen Thi Thu: Methodology, Writing – review and editting
Vu Le Hoang: Methodology, Writing – original draft
Duy Tran Ngoc Bao: Investigation, Methodology
Vu Tran Pham: Conceptualiation, Investigation, Methodology
Dung Ta Quoc: Conceptualiation, Investigation

REFERENCES

1. McCormack MD. Neural computing in geophysics. The Leading Edge. 1991;10(1):11–15. Available from: https://doi.org/10.1190/1.1436771.
2. Rezaee MR, et al. Intelligent approaches for the synthesis of petrophysical logs. Journal of Geophysics and Engineering. 2008;5(1):12–26. Available from: https://doi.org/10.1088/1742-2132/5/1/002.
3. Ghavami F. Developing synthetic logs using artificial neural network: Application to Knox County in Kentucky. West Virginia University Libraries. 2011;.
4. Long W, et al. Pseudo density log generation using artificial neural network. In SPE Western Regional Meeting. Society of Petroleum Engineers. 2016;Available from: https://doi.org/10.2118/180439-MS.
5. Salehi MM, et al. Estimation of the non records logs from existing logs using artificial neural networks. Egyptian Journal of Petroleum. 2017;26(4):957–968. Available from: https://doi.org/10.1016/j.ejpe.2016.11.002.
6. Yann L, Yoshua B, Geoffrey H. Deep Learning. 2015;.
7. Chu MG, et al. Determination of an infill well placement using a data-driven multi-modal convolutional neural network. Journal of Petroleum Science and Engineering. 2020;195:106805. Available from: https://doi.org/10.1016/j.petrol.2019.106805.
8. Specht DF. Probabilistic neural networks. Neural Networks. 1990;3(1):109–118. Available from: https://doi.org/10.1016/0893-6080(90)90049-Q.
9. Khan S, et al. A Guide to Convolutional Neural Networks for Computer Vision, Morgan & Claypool. 2018;Available from: https://doi.org/10.2200/S00822ED1V01Y201712COV015.
10. Goodfellow I, et al. Deep Learning. MIT Press. 2016;.
11. Sewak M, et al. Practical convolutional neural networks: Implement advanced deep learning models using python, Packt Publishing. 2018;.