P-Wave Sonic Log Predictive Modeling with Optimal Artificial Neural Networks Topology

Labiba M. Alhelfi[a],*, Hana M. Ali[b], Semaa H. Ahmed [c]

[a] Department of Mathematics, College of Sciences, University of Basrah, Basrah 61004, Iraq . *Email: dhmary033@gmail.com.*

[b] Department of Mathematics, College of Sciences, University of Basrah, Basrah 61004, Iraq . *Email: hana.ali@uobasrah.edu.iq.*

"c Department of Petroleum Engineering, College of Engineering, University of Basrah, Basrah 61004, Iraq" . *Email: semaaalessa@gmail.com.*

A R T I C L E I N F O

A B S T R A C T

Given the financial challenges facing the oil and gas industry, the value of the information is considered relatively high; therefore, data science has been an alternative compensating tool. This study aimed to find an optimal neural network topology that provides an ideal data solution by studying neural network topology. Therefore, we trained different neural networks topologies in terms of the number of hidden neurons and layers. Volve oil field data is used in this study to predict the compressional sonic wave travel time. Optimal Neural Network topology found using five hidden layers and five hidden neurons while using a single layer with different numbers of hidden neurons was ineffective. The highest training and testing accuracy with a single hidden layer found 0.94 and 0.914, respectively. In contrast, it was found 0.947 and 0.934 with 50 hidden neurons and five hidden layers. Yet, increasing the number of hidden layers and hidden neurons is found to cause overfitting; therefore, only an optimal topology is a critical factor.

MSC. 41A25; 41A35; 41A36,,

1. Introduction

For centuries, petroleum has been the world's most significant energy source; however, oil and gas prices fluctuations adversely affected capital investment. Therefore, the petroleum industry is shifting towards the fourth paradigm shift, which is data science [1]. Data science engagement has a significant positive impact on economics, that was adversely affected by the high value of information due to market demand challenges.

One of the most critical parameters in the oil industry is the compressional delta-t travel time (P-wave sonic log). The last is used to obtain effective porosity, water saturation, lithology, and seismic correlation [2]. However, obtaining this information is subjected to different challenges, such as the wellbore conditions and the value of information [2]. Therefore, artificial intelligence (AI) played a significant role in generating synthetic data with least capital and time investment.

∗Corresponding author:- *Labiba M. Alhelfi*

Email addresses:- *dhmary033@gmail.com*

Communicated by: Dr. Rana Jumaa Surayh aljanabi.

Artificial neural networks (ANNs) have been applied in several industrial fields [3, 4, 5], including the oil and gas industry, due to their ability to represent the high dimensional data to discover the complex non-linear features between inputs and the desired output such as: real-time quality monitoring in the oil refinery [6], logging predictions [7], inter-well connectivity between producers and injectors [8], drilling problems [9], etc.

Multi-layer perceptron (MLP) is one of the fundamental types in artificial neural networks. MLP has three or more artificial layers of neurons (an input layer, hidden layer/layers, and an output layer) [10, 11]. Where neurons are connected by synapsis and the strength of the connection is defined by the synaptic weight. A high accuracy MLP model is found with the optimum weights. The last is optimized by the process of Backpropagation, which attempts to minimize the Mean Square Error (MSE).

MLP is a regression analysis tool that effectively can approximating functions in high-dimensional spaces. This ability come from the nonlinearity embedded in the neurons. Furthermore, Unlike the conventional linear regression algorithms, MLPs do not suffer from dependencies between the input dimensionality and the convergence rate [12]. In this paper, a Deep MLP regression, ANN with multiple layers, has been used to build a P-wave sonic log regression model considering high dimensional input of different well logs, which are: Gama Ray (GR), Caliper (CALI), Resistivity(RT), Neutron Porosity (NPHI), Bulk Density (RHOB), Density correction (DRHO), Measured Depth (MD), True Vertical Depth (TVD), and Photoelectric factor (PEF).

2. Related Work

Recently, ANN had a remarkable ability to determine nonlinearities between inputs and outputs and effectively predict various real-world properties. In [13], an ANN was applied to predict the P-wave sonic logs based on three logs (Porosity, Density, Gamma-ray). The authors compared their model with the multiple linear regression model MLR, and they obtained an excellent output.

In addition, a type of feedforward ANN, called a supervised Probability Neural Network (PNN), was utilized to predict the P-wave logs based on three basic well logs: Density Gamma-ray, and Resistivity logs [14]. Those three logs were used as inputs vectors to predict the P-wave log. The authors claimed that using PNN gave the best accuracy after comparing it with the traditional methods such as Gardner and Faust method.

Besides P-wave logs, other authors tested several shear sonic wave travel time (S-wave sonic log) models: including support vector machine, regression tree, linear regression, and gaussian process regression. And the Gaussian process regression performed better than the simple linear regression model [15].

Furthermore, different S-wave sonic log prediction models were deployed based on five other machine learning methods [16]: linear regression, random forest, support vector machine, and XGBoost, and ANN. The result concluded that XGBoost surpassed the rest of the models in predicting the S-wave logs.

In this paper, MLP is used for P-wave sonic log prediction. Our model has statistically been built based on nine input features: Gama Ray, Caliper, Resistivity, Neutron Porosity, Bulk Density, Density correction, Measured Depth, True Vertical Depth, and Photoelectric factor as shown in **Fig. 2**. Furthermore, to obtain a better correlation coefficient value, we have tested different MLP topologies.

3. Methodology

Descriptive and predictive models were performed in series after the data mining process. The descriptive analysis enhanced the predictive model by utilizing the key performance features as an input. Sequentially, the predictive models were generated later for different ANN topologies.

In 2018, a vast amount of data of Volve field was disclosed to the public; thus, it was an excellent opportunity for researchers, IT companies, and universities to utilize this data to deal with several oil issues. The Volve field is located in the Norwegian North Sea, with a field life of 2008-2016, produced 63 million barrels from the Hugin Formation [17].

The Volve field is a small dome-shaped structure and was formed by the collapse of adjacent salt ridges during the Middle Jurassic age. Faults as a consequence of salt tectonics are the dominant structures in this field. The structure is heavily faulted, especially in the western area, most probably caused by the influence of regional extension. The thickness of the reservoir varied as of 20 m at the top of the crest to 100 m on the structure flanks [18].

Since this study aims to find the best topology of the P-sonic log ANN predictive model, it is important to note that the input and the output parameters shares a hidden physics. Therefore, ANN is trained to understand the hidden physics in a complex, vast amount of data. At the same time, the degree of ANN topology complexity is to be linked to the data physics complexity.

3.1 Dataset Mining

In this section, data was cleaned by removing the well logs available for less than ten wells and dropping wells with less than 12 available logs, along with dropping rows with empty records.

Table (1) shows the seven wells' data count within the dataset plus the available well logs. We utilized the second well as testing well for later evaluation of our model, and thus, it has not been embedded with the training dataset.

TABLE 1: NEURAL NETWORKS DATA INPUT SUMMARY

Wells	Rows Count	Logs
'15/9-F-1'	9822	'MD (M)'
'15/9-F-1 B' (Testing)	3262	'TVD (M)'
'15/9-F-11 A'	11374	'CALI (inches)'
'15/9-F-11 T2'	19433	'DRHO (g/cm3)'
'15/9-F-15'	7146	'DT (us/ft)'
'15/9-F-15 A'	9361	'GR (API)'
		'NPHI (v/v)'
		'RHOB (g/cm3)'
'15/9-F-1 A'	10222	'RT (ohm.m)'
		'PEF (b/elec)'

On the other hand, the dataset contains different physical features with various scales, e.g., "MD" values were found within thousands while the "NPHI" is on decimals value (refer to **Fig. 2**). Thus, before determining the inputs and output for our model, the dataset must be transformed. Therefore, we used the Yeo Johnson Power Transformation method, whereas the power parameter lambda (λ) is changed until the best approximation of a Gaussian likely distribution is reached [19].

$$\psi(\lambda, y) \begin{cases} ((y+1)^\lambda - 1)/\lambda & if\ \lambda \neq 0, y \geq 0 \\ \log(y+1) & if\ \lambda = 0, y \geq 0 \\ -[(-y+1)^{2-\lambda} - 1)] / (2-\lambda) & if\ \lambda \neq 2, y < 0 \\ -\log(-y+1) & if\ \lambda = 2, y < 0 \end{cases}$$

Regarding the data preparation, outliers of our cleaned datasets were removed using the one-class support vectors machine [20]; thus, 52 outliers were removed from the dataset. According to [21], One-Class SVM distances all the data points from their origin, maximizing the distance from this hyperplane. Then it defines the data's density probability across the hyperplane (returns +1 for the included data within the hyperplane and −1 for data located outside). The normalized data made feature extraction easier, especially for the given well logs: "NPHI" and "RT" which were hard to distinguish without data transformation. Once data is normalized, the data is partitioned in the following fashion:
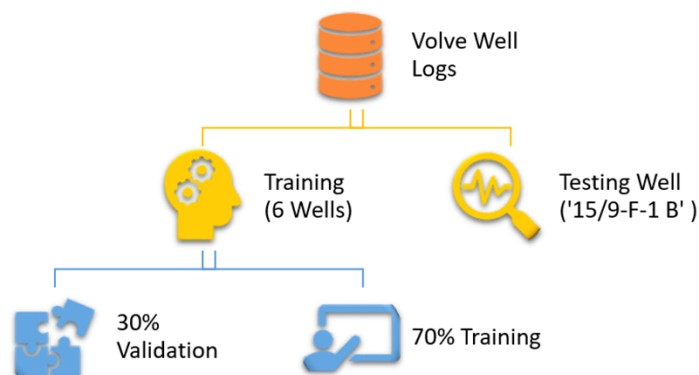

Fig. 1: DATA PARTITIONING SCHEMATIC

As mentioned earlier, only "15/9-F-1 B" was set aside for testing, which will not contribute to the MLP training, but to test the final model for efficiency of generalization. The rest of the wells shown in Table 1 were used to build the predictive MLP model with 70% of data only, whereas 30% is left for model validation. The validation will play 'the watching dog role'; it will only help decide when to stop the training. Since the designed MLP model is intended to predict the P-wave sonic log (or DT), the output layer will only include the P-wave sonic log, whereas the rest represent the input layers.

3.2 MLP Architecture

In this section, we defined our MLP model and the initialized hyperparameters. It is well known that there is no general and explicit method for choosing MLP hyperparameters. Thus, different MLP hyperparameters can help to satisfy the global maximum. Nevertheless, only three hyperparameters were tested. We tested three activation functions: "Logistic", "Tanh", and "Relu". Moreover, we tried four maximum iterations: 200, 800, 1200, and 1800. Our validation fraction has three values: 10%, 20%, and 30%. And "Adam" optimization algorithm is used for decreasing the objective function of error.

On the other hand, as the ANN topology (number of hidden neurons and hidden layers) plays an important role in the learning phase, any increase or decrease in the number of hidden neurons/layers significantly affects the entire learning process. It changes the number of weights and the amount of nonlinearity embedded in the activation functions, thus affecting the prediction accuracy. According to novel research, the complexity of data, in terms of the number of clusters, could give a hint about the optimal topology; the more clusters in a dataset, the more neurons/layers it requires to reach an acceptable error. Thus, we have exploited the hierarchical cluster analysis to explore the situation in the data topology in terms of clusters to see such a relation between the data complexity and the optimal topology (hidden neurons/layers). The tested ANN topologies were the total possible combinations of the following vectors: Neurons = [10:100] and Layers = [1:10]. Results were later compared at each topology to detect the optimal topology that can successfully express the complexity of our dataset.
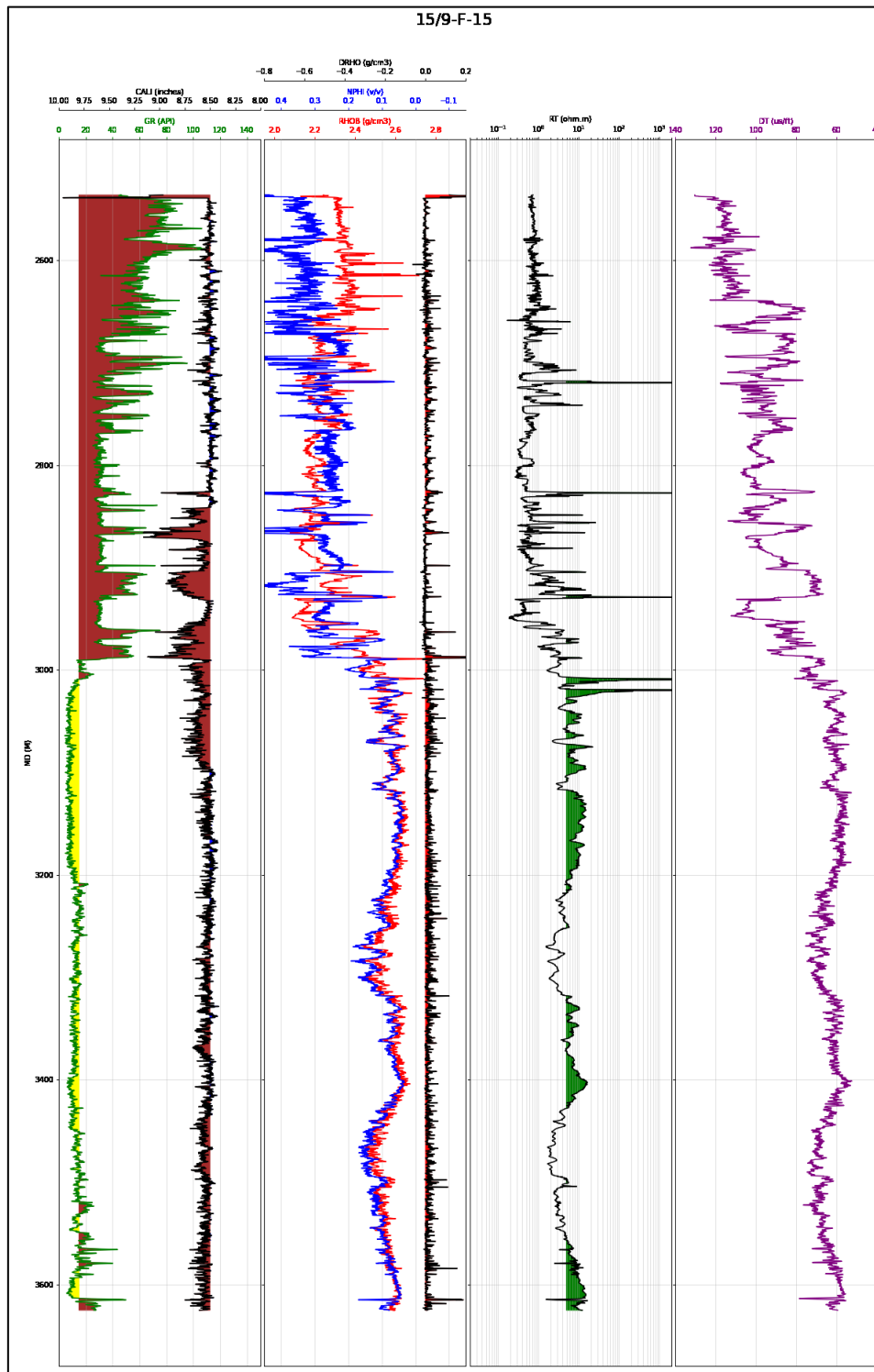
Fig. 2: Single well dataset example after the data mining process

4. RESULTS and DISCUSSION

   This section shows the results of the MLP predictive model with eight inputs (or features) and one output, trained on six wells.

4.1 Data Complexity

   As we mentioned earlier, it is essential to understand the dataset before building the ANN model, as data complexity will impact the training process's hyperparameters selection. **Fig. 3** shows a Hierarchical cluster analysis of the dataset. The clusters were realistic as the depth logs were set within the same cluster (MD and TVD) and for the wellbore geometry-dependent well logs (DRHO and CALI). We noted four groups with relatively similar features; each group has two physically similar features (the distance between the two features is small). It is important to note that "GR" and "NPHI" have large distances in compared to any other physical features, and this means having a spacing between some input which will indicate the data complexity and features interactions because the network will struggle to find the relationship between the inputs due to the presence of this spacing. The network will likely need more than one hidden layer and many hidden neurons to express such complexity.
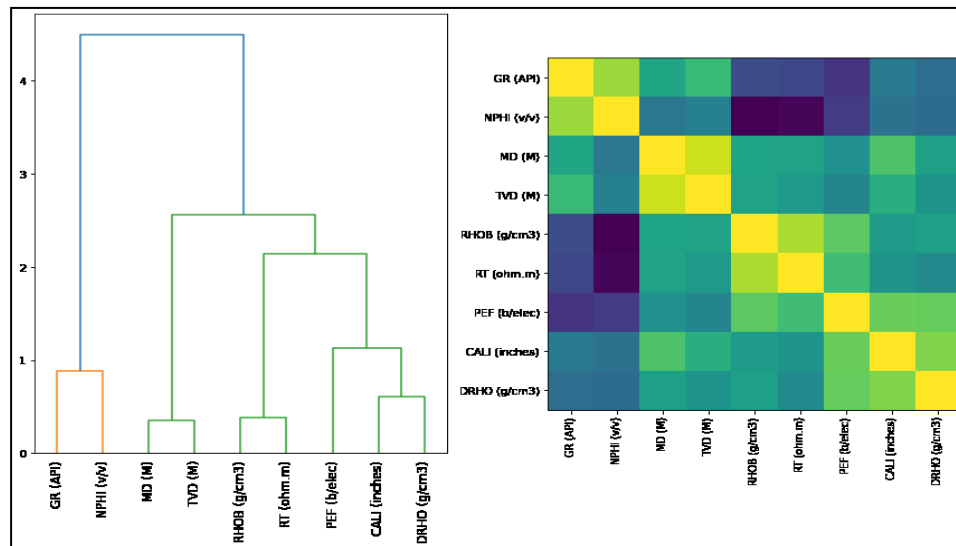

Fig. 3: Hierarchical cluster analysis of the well log data (Dataset Complexity)

4.2 Optimal Topology and Hyperparameters

   Once the potential hyperparameters are tested, the optimal MLP predictive model hyperparameters for the given dataset were: 200 iterations, 0.0002 Alpha, Relu activation function, and 30% validation fraction. Different Hidden Neurons and Hidden layers were tested with [10:100] and [1:10], respectively, after obtaining the hyperparameters and understanding the training data's physical complexity. **Fig. 4** shows the result determining the training and testing error for neurons number versus layers number.
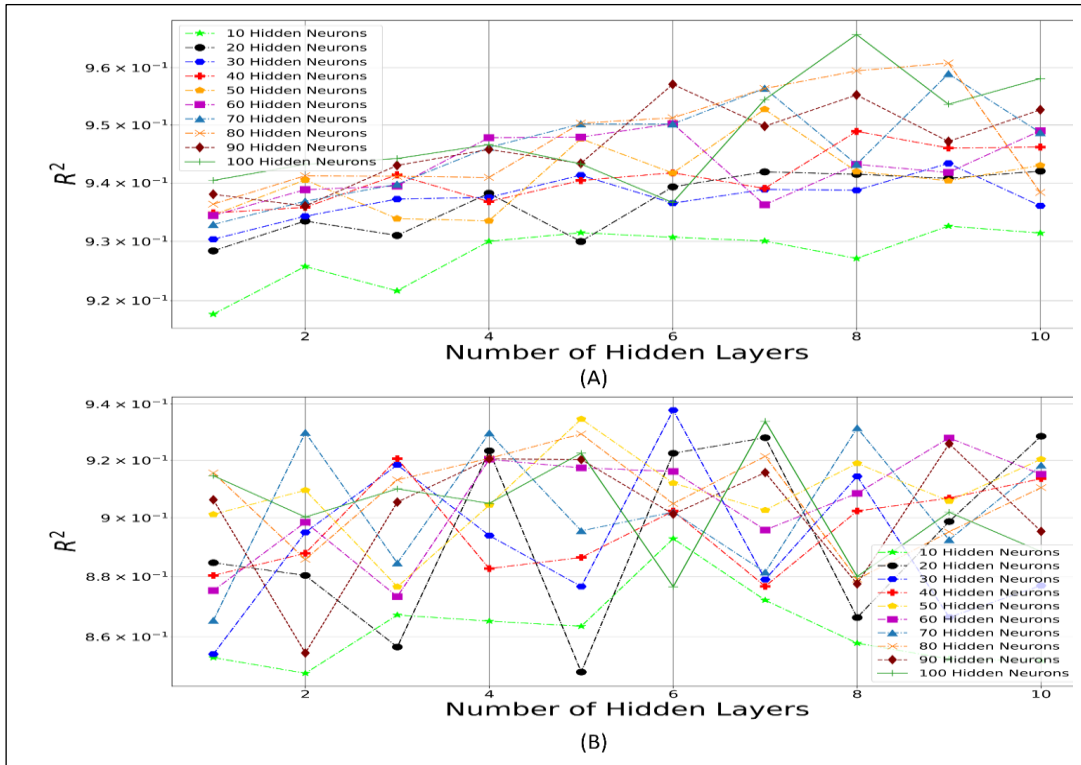
Fig. 4. R^2 of No. of Hidden Layers Versus No. of Neurons for (A): Training Error. (B) Testing Error.

Fig. 4 shows R^2 results for the training and testing dataset based on different predictive models' topologies. It shows a clear example for improving training as a function of the number of layers and number of neurons increase; nevertheless, high training accuracy does not necessarily indicate a good generalization due to the overfitting. The last may exhibit noise generalization on the testing dataset. According to the empirical analysis, the best topology is within fifty hidden neurons and five hidden layers. The training accuracy is high, and there is no gap between the training accuracy and the testing accuracy.

The optimal topology, five layers each with 50 neurons, was chosen based on the highest testing R^2 obtained from testing different 100 MLP topologies. The results were classified as well-trained, overtrained, and poorly-trained topologies, as shown in Table 2, proving that the high training accuracy of 0.9658 gave relatively low testing accuracy (classified overtrained). The well-trained models agreed with the proposed complexity; several hidden layers satisfied the model objectives of high testing accuracy. These findings shed light on the complex hidden physics that need more neurons and more hidden layers for more weight fractions (granularity increase). On the contrary, no trained topologies could reach a good accuracy with a single layer regardless of the given number of hidden neurons values. Therefore, we have compared multiple topologies for the well-trained, overtrained model and poorly-trained topologies and the best single-layer models.

TABLE 2: MLP PREDICTIVE MODELS TOPOLOGIES VERSUS R² SCORE

Topology Status	R² Score for Training	R² Score for Testing	Number of Hidden Layers	Number of Hidden Neurons
Well-trained	0.9365	0.9378	6	30
	0.9476	0.9347	5	50
	0.9544	0.9339	7	100
	0.9432	0.9315	8	70
Overtrained	0.9659	0.8800	8	100
	0.9609	0.8952	9	80
	0.9595	0.8791	8	80
	0.9590	0.8925	9	70
Poorly-trained Model	0.9258	0.8481	2	10
	0.9300	0.8485	5	20
	0.9314	0.8522	10	10
	0.9326	0.8528	9	10
Best Single Layer Topologies	0.9364	0.9156	1	80
	0.9404	0.9146	1	100

4.2.1 Well-Trained Topologies

Fig. 5 shows a comparison of both the actual P-wave sonic log versus the predictive model P-wave sonic log. The comparison is shown for both training and testing wells as a function of the measured depth. Both Fig. 5 and Fig. 6 presented an identical match that analytically validates the P-wave sonic log predictive model. Yet, 50 hidden neurons for five layers were the best chosen topology for this study.
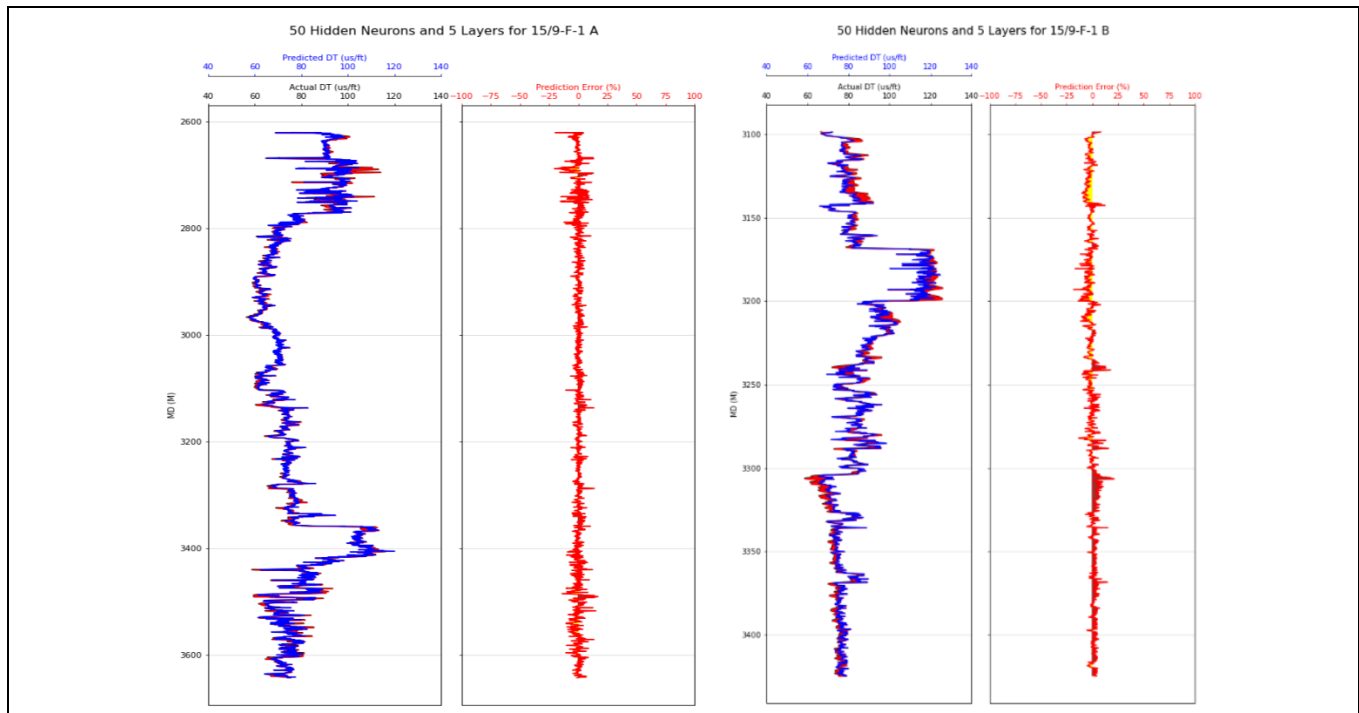

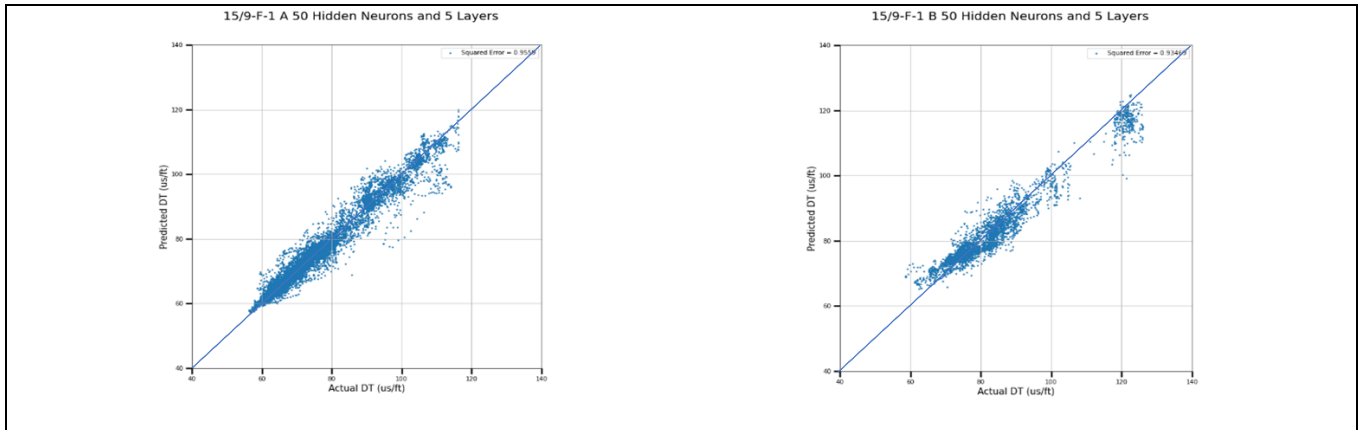Fig. 5. A comparison of well-trained topologies with 50 hidden neurons and five hidden layers

Fig. 6. R^2 for '15/9-F-1 B' and '15/9-F-1 A' with 50 hidden neurons and five hidden layers

4.2.2 OVER TRAINED TOPOLOGY

Fig. 7; using the 80 hidden neurons and the eight layers could cause noise generalization problems as shown in **Fig. 8** as data scattered off the straight-line trend of '15/9-F-1 B'.
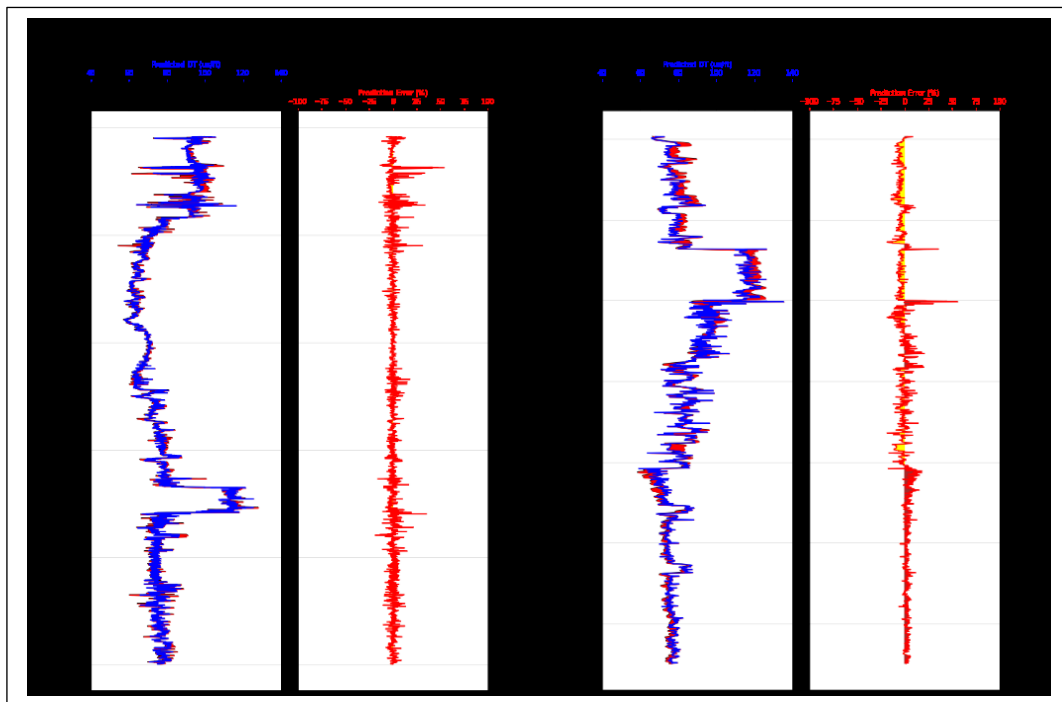

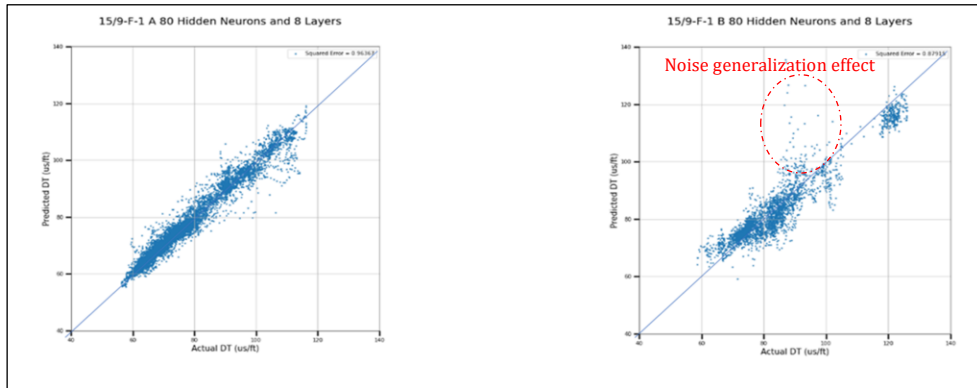Fig. 7. An example of overtrained topology with 80 hidden neurons and eight hidden layers

Fig. 8. R^2 for '15/9-F-1 B' and '15/9-F-1 A' with 80 hidden neurons and eight hidden layers

4.2.3 POORLY TRAINED TOPOLOGY

The given topologies for this class didn't help to reveal the data complexity for the testing well. The predicted and the actual P-wave sonic log did not match relatively, causing some error (refer to **Fig. 9** and **Fig. 10**).
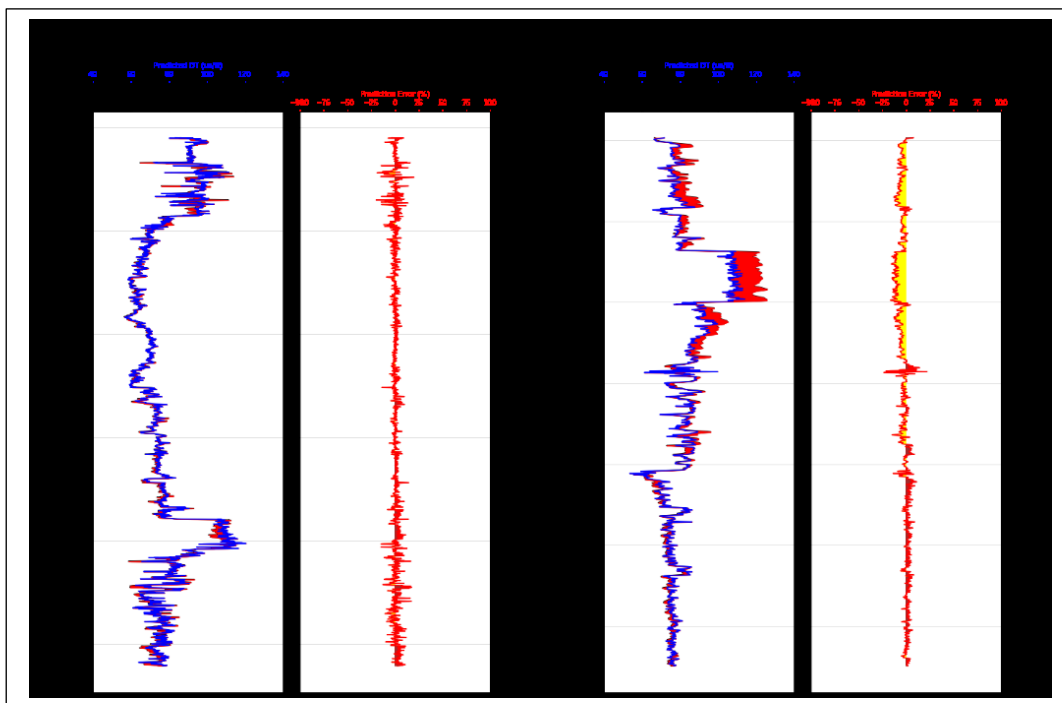

Fig. 9. An example of poorly trained topology with ten hidden neurons and two hidden layers
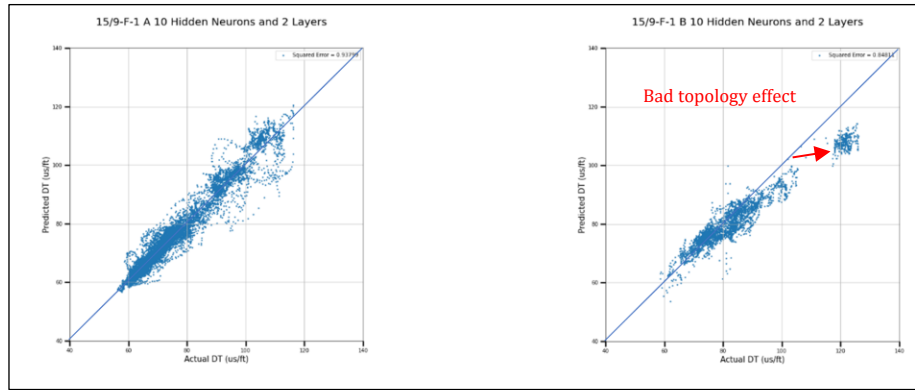
Fig. 10. R^2 for '15/9-F-1 B' and '15/9-F-1 A' with ten hidden neurons and two hidden layers

4.2.4 SINGLE LAYER TOPOLOGY

It was noted in **Fig. 11** and **Fig. 12** that having a single layer for all the given hidden neurons did not help to improve the model training based on Table 2. Nevertheless, the highest accuracy for the testing dataset was obtained as 0.915 with one hidden layer and 80 hidden neurons. At the same time, the other topology had 100 hidden neurons which notates that the complex data might require a high number of hidden neurons due to the granularity increase
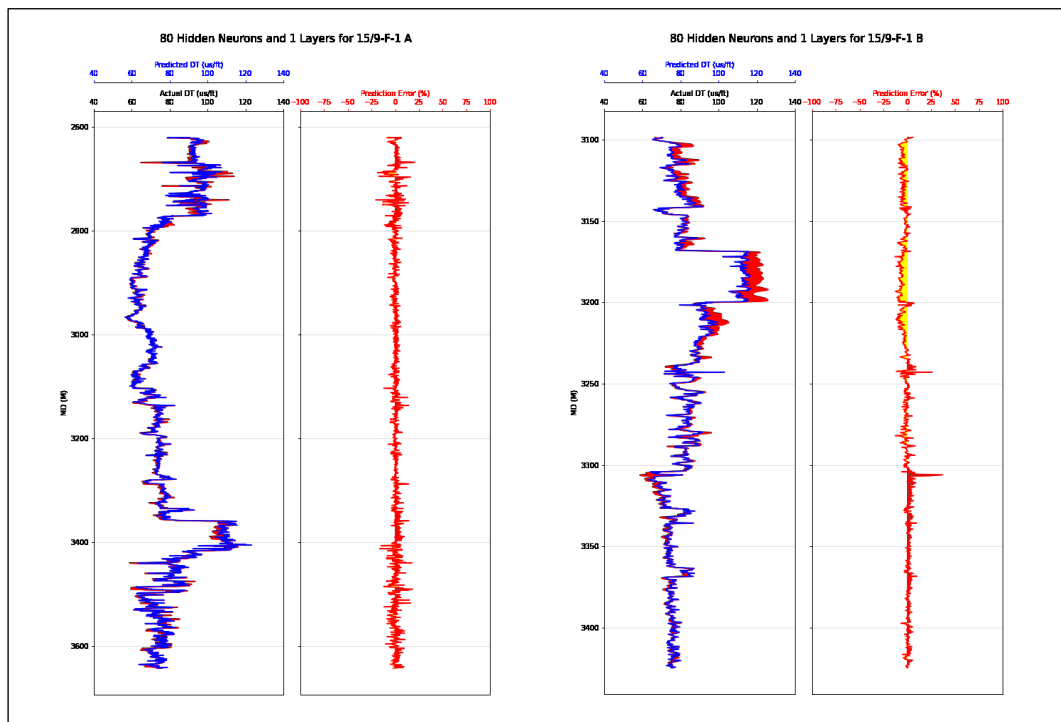

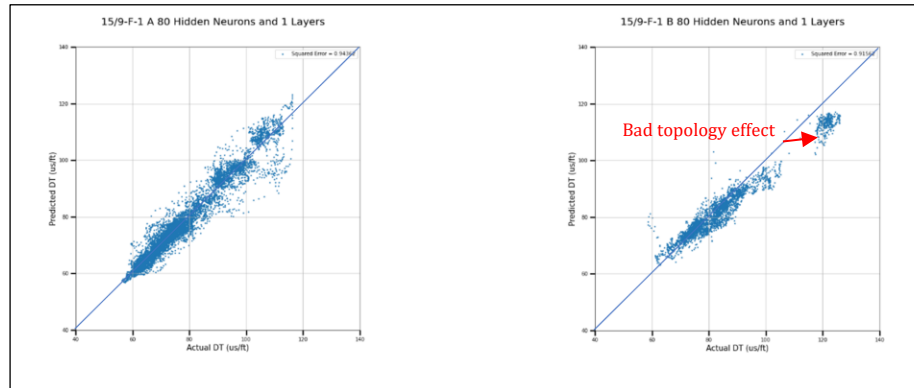Fig. 11. An example of single layer topology with 80 hidden neurons

Fig. 12. R2 for '15/9-F-1 B' and '15/9-F-1 A' with 80 hidden neurons and one hidden layer

5. CONCLUSION

This study intended to choose the best topology for neural networks to predict the P-wave sonic log. The study tested different 100 topology sets, and Hierarchical cluster analysis showed multiple clusters indicating some feature complexities. The research also showed that building the predictive model with one hidden layer didn't improve the model compared to the other topologies. Accordingly, it was recommended to increase the number of the hidden neurons as the data showed high complexity. In addition, deploying the predictive model with multiple hidden layers and hidden neurons might cause overtraining. Therefore, understanding the dataset topology versus ANN topology is essential for an optimal P-sonic log predictive model. As a result, it was concluded that the best topology is found with 50 hidden neurons and five hidden layers.

Acknowledgments

References

[1]   Gressling, Thorsten. "2 Data science: the "fourth paradigm" of science." In Data Science in Chemistry, pp. 4-7. De Gruyter, 2020.

[2]   Liu, Hongqi. Principles and applications of well logging. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017.

[3]   Habib Al- Sharoot, M., & Yousif Abdoon, E. (2017). Prediction by using Artificial Neural Networks and Box-Jenkins methodologies: Comparison Study. *Journal of Al-Qadisiyah for Computer Science and Mathematics*, *9*(2), Stat Page 1 - 16. https://doi.org/10.29304/jqcm.2017.9.2.325

[4]   H. AL - Sharoot, M., & A. Saheb, D. (2017). Forecasting of electricity power consumption in AL - Qadisiya province by using time series models & Artificial neural networks models. *Journal of Al-Qadisiyah for Computer Science and Mathematics*, *6*(1), 9-26. Retrieved from https://qu.edu.iq/journalcm/index.php/journalcm/article/view/132

[5]   Habib Al- Sharoot, M., & Yousif Abdoon, E. (2017). Prediction by using Artificial Neural Networks and Box-Jenkins methodologies: Comparison Study. *Journal of Al-Qadisiyah for Computer Science and Mathematics*, *9*(2), Stat Page 1 - 16. https://doi.org/10.29304/jqcm.2017.9.2.325

[6]   H. Singh, A. K. Pani, and H. K. Mohanta, "Quality monitoring in petroleum refinery with regression neural network: Improving prediction accuracy with appropriate design of training set," Measurement, vol. 134, pp. 698-709, 2019.

[7]   M. Sabah, M. Talebkeikhah, D. A. Wood, R. Khosravanian, M. Anemangely, and A. Younesi, "A machine learning approach to predict drilling rate using petrophysical and mud logging data," Earth Science Informatics, vol. 12, no. 3, pp. 319-339, 2019.

[8]   H. Cheng, X. Han, P. Zeng, H. Yu, E. Osipov, and V. Vyatkin, "ANN based Interwell Connectivity Analysis in Cyber-Physical Petroleum Systems," in 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), 2019, vol. 1: IEEE, pp. 199-205.

[9]   Abduljabbar Saad, I. (2018). An Efficient Classification Algorithms for Image Retrieval Based Color and Texture Features. Journal of Al-Qadisiyah for Computer Science and Mathematics, 10(1), Comp Page 42 - 53. https://doi.org/10.29304/jqcm.2018.10.1.350

[10]   Mohammed, Y., & Saleh, E. (2020). Investigating the Applicability of Logistic Regression and Artificial Neural Networks in Predicting Breast Cancer. Journal of Al-Qadisiyah for Computer Science and Mathematics, 12(2), Math Page 63-. https://doi.org/10.29304/jqcm.2020.12.2.697

[11]   Menzies, Tim, Ekrem Kocaguneli, Burak Turhan, Leandro Minku, and Fayola Peters. Sharing data and models in software engineering. Morgan Kaufmann, 2014.

[12]   K.-L. Du and M. Swamy, "Elements of computational learning theory," in Neural networks and statistical learning: Springer, 2019, pp. 65-79.

[13]   M. Zoveidavianpoor, A. Samsuri, and S. R. Shadizadeh, "Prediction of compressional wave velocity by an artificial neural network using some conventional well logs in a carbonate reservoir," Journal of Geophysics and Engineering, vol. 10, no. 4, p. 045014, 2013.

[14]   A. Ogbamikhumi, S. Salami, and W. Uwadiae, "Neural Network Prediction of P-wave Log for Reservoir Characterization in the Niger Delta Basin," Nigerian Journal of Technological Development, vol. 17, no. 1, pp. 28-32, 2020.

[15]   I. Bukar, M. Adamu, and U. Hassan, "A machine learning approach to shear sonic log prediction," in SPE Nigeria Annual International Conference and Exhibition, 2019: OnePetro.

[16]   S. Liu, Y. Zhao, and Z. Wang, "Artificial Intelligence Method for Shear Wave Travel Time Prediction considering Reservoir Geological Continuity," Mathematical Problems in Engineering, vol. 2021, 2021.

[17]   Bhamidipati, Subrahmanya Chandra. "Power generation from waste heat during oil production." PhD diss., Missouri University of Science and Technology, 2021.

[18]   Johnston, David H., ed. Methods and applications in reservoir geophysics. Society of Exploration Geophysicists, 2010.

[19]   S. Weisberg, "Yeo-Johnson power transformations," Department of Applied Statistics, University of Minnesota. Retrieved June, vol. 1, p. 2003, 2001.

[20]   Dawson, C. (2018, November). Outlier Detection with One-Class SVMs. Retrieved from towards data science: https://towardsdatascience.com/outlier-detection-with-one-class-svms-5403a1a1878c.

[21]   B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support vector method for novelty detection," in NIPS, 1999, vol. 12: Citeseer, pp. 582-588.