

The Nigerian Association of Mathematical Physics

Journal homepage: <https://nampjournals.org.ng>



Ensemble Neural Network Approach for Predicting Water Saturation Using Well Logs

¹Onwusinkwue S., ¹John O. J., ²Agbonota G. O.,

¹Department of Physics, University of Benin, Benin City, Nigeria.

² Faculty of science lab technology, University of Benin, Benin City, Nigeria.

ARTICLE INFO

Article history:

Received xxxxx

Revised xxxxx

Accepted xxxxx

Available online xxxxx

Keywords:

Water saturation, ensemble model, neural network, well log, MATLAB, core data, learning, reservoir

ABSTRACT

Accurate estimation of water saturation (S_w) is vital for hydrocarbon recovery from a reservoir. Empirical equations (EE) such as: Archie, Simandoux, Waxman-Smith and Dual water model are used to estimate water saturation. The results are unreliable and strongly depend on costly and time-consuming core analysis. This study presents ensemble neural network approach for predicting water saturation using well logs. The results predicted by the ensemble model in MATLAB was evaluated against (i) core data (ii) empirically derived data (iii) single expert network model, using standard performance metrics such as MSE, coefficient of determination (R^2) and coefficient of regression (R). The results show that the ensemble approach significantly improves prediction accuracy because it closely fits the core data.

1. Introduction

The pore spaces in reservoir rocks are occupied by fluids-usually water and hydrocarbons. The relative volumes of water and hydrocarbons in the pore volume of the reservoir rock are designated as saturations. Water saturation (S_w) is the fraction of the pore volume occupied by water. It is a vital parameter to estimate the hydrocarbon saturation in the formation ($1-S_w$). [1] Accurate determination of water saturation is important for better reservoir characterization and subsequent recovery of oil from hydrocarbon fields. Some of the common methods employed to determine the water saturation are petro-physical log evaluation, saturation–height function from capillary pressure data and core analysis. [2]. Estimating water saturation is one of the most challenging petro-physical calculations.

*Corresponding author: Onwusinkwue S.

E-mail address: shedrack.onwusinkwue@uniben.edu

DOI: <https://doi.org/10.60787/jnamp.vol73no.754>

1118-4388 © 2024 JNAMP. All rights reserved

Special Core Analysis (SCAL) is the most reliable direct method to determine water saturation, the main drawback with this method is the high cost for the coring process and the extensive laboratory analysis involved. Water saturation is often estimated from well logs using empirical equations. [3] These equations are complex and prone to errors because; they are based on assumptions, adjustment of parameters depending on location and cross plot between logs. [4]

Reservoir properties, such as porosity, water saturation and permeability can be obtained from the processing and interpretation of well-logs. Modern reservoir characterization has evolved from the use of experts' knowledge and statistics in the analyses of well-logs to the use of artificial intelligence (AI). [5, 6] This study proposes to use ensemble neural network to predict water saturation (S_w) from the conventional well-logs and to compare the results obtained with values available from core analysis and that calculated from empirical equation.

2. NEURAL NETWORK

Neural Network is a computational model inspired by the structure and functioning of neurons in the human brain. It is made up of interconnected nodes called artificial neurons, organized in layers. Basically, neural networks consist of an input terminal, one or more hidden layers, and an output layer. Each layer contains one or multiple neurons, each neuron has an activation function, bias and connections to other neurons. Inter-connection between neurons carry numerical data, called weights. Since neural network is designed to mimic biological learning processes as it occurs in human brains, it has the capacity to learn complex relationships between data when trained. [7]

Figure 1 is a typical neural network configuration showing the interconnection of artificial neurons organized in layers: input layer (i), one or more hidden layers (h) and output layer (o)

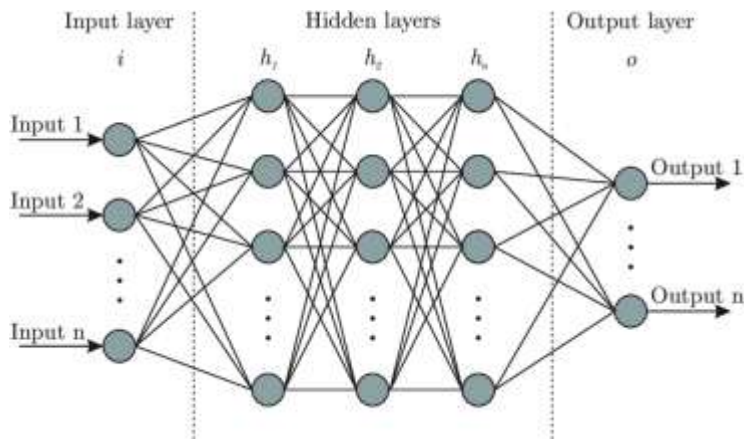


Figure 1: A typical neural network architecture

During training, the training algorithm minimizes the mean square error (MSE) between the predicted outputs of the network and observation outputs used in the training dataset. The learning occurs as the network continues to adjust and update the weights and bias of the neurons, in its search to provide the optimal weight/bias configuration that maps the relationship between the input-output data. A well-trained neural network with the updated best weights can generalize for unseen input data. The three learning algorithm used in neural networks are: supervised, unsupervised and reinforced learning.

Generally, the relationship between the input and output datasets in a neural network model is given by Eq. 1

$$y_k = f_0 \left[\sum_j w_{kj} f_h (w_{ji} x_i + b_j) + b_k \right] \quad (1)$$

Where: x_i represents an input vector; w_{ji} is connecting weights from the i th neuron in the input layer to the j th neuron in the hidden layer and w_{kj} is the connecting weights from j th neuron in the hidden

layer to the k th neuron in the output layer; b_j and b_k are the biases of j th hidden neuron and the k th output neuron, f_h and f_o denote the activation functions for the hidden and output neuron, respectively. Feed-Forward Multilayer Perceptron (MLP) trained with back-propagation algorithm is the most common type of neural network deployed in petroleum engineering applications. [5, 8] An Ensemble Network (EN) is made up of two or more candidate networks, often called expert networks that are joined together. Usually these expert networks have different network architecture, and they are trained differently on same data set before combining their outputs. The expectation is that the differently trained experts converge to different local minima on the error surface, and such that their overall performance is improved by combining their outputs by ensemble averaging. [9]. Figure 1 is typical ensemble neural network consisting of four (4) expert networks.

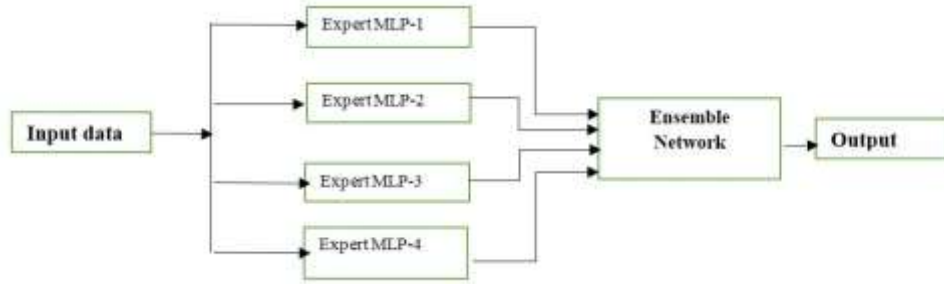


Figure 1: Ensemble Neural Network

The output of the ensemble network is given as:

$$y_{EN} = \frac{1}{N} \sum_{k=1}^N y_k \quad (2)$$

Where: y_k is the output of each expert network. And N is the no of expert networks.

Data driven method like neural network has become popular in the domain of reservoir characterization due to its ability to map complex non-linear relationships between data. By training a neural network on a dataset of well logs and corresponding core measurements, the networks can learn complex relationships between log responses and rock properties, and can make accurate predictions in new wells or areas where core data is unavailable.[14].

Okon et al developed a artificial neural network (ANN) model using MATLAB 2015 neural fitting toolbox. They deployed the ANN to successfully predicted porosity, permeability and water saturation using logging depth interval, density log, resistivity log and gamma log. The porosity, permeability and water saturation obtained in their study for Niger Delta region showed higher correlation value with available core data[5]

Ahmed et al designed Artificial neural networks (ANNs) and the Adaptive Neuro-Fuzzy Inference System (ANFIS) to estimate water saturation, then compared their result to water saturation obtained using empirical equation as: Archie, Simandoux, Waxman-Smith, Dual water model and Indonesian equations. Their result proved that the water saturation estimated by the ANN and ANFIS models displayed better fit to core data compared to results from empirical equations. [3]

In order to improve the performance of data driven methods, researchers have embrace the idea of combining multiple network systems. When creating ensemble network, the expectation is that the differently trained experts converge to different local minima on the error surface, and the overall performance is improved by combining their outputs in some way. [11] Ahmadi and Chen 2019, used ensemble network consisting of multiple layer perceptron (MLP), genetic algorithm (GA), fuzzy decision tree (FDT), imperialist competitive algorithm (ICA) and particle swarm optimization (PSO) to predict porosity and permeability. The result of their study showed that the porosity and permeability values

predicted by the ensemble network was better when compared to the prediction obtained for the ensemble members. [12]

Sun and Dong deployed a committee network consisting of radial basis feedforward network (RBFNN) and genetic algorithm (GA) to predict permeability. According to their study, the RBFNN-GA had higher accuracy in permeability prediction compared to the candidate RBF or GA model. [13] Sambo et al. integrated of artificial neural network (ANN) and NeuroFuzzy System(NFS) with seismic inversion to predict water saturation in wells in Malaysian hydrocarbon field. Their study showed that incorporation of AI technology in water saturation prediction workflow from pre-stack seismic data resulted in superior accuracy and better performance. [14]

3. METHODOLOGY

3.1 Data Description

The well logs and core data used in this study was sourced from Shell Petroleum Development Company (SPDC) in Port-Harcourt, Nigeria. The data was obtained from two (2) wells (Well-A and Well-B). The original data had 1200 sample points for the well logs and 188 sample points for core data. The wide gap between log and core data is due to the fact cores are acquired at choice intervals within the borehole depth.

The wells had calculated/derived values for water saturation. This values will be treated as calculated water saturation using empirical equations (EE). The petro-physical interpretation for the wells clearly delineated regions of shale, oil, gas and water. This interpretation was used to carry out log-core calibration so that that training, validation and testing data were representative of the formation.

Well-A data was used for training and Well-B data was used to test or simulate the trained networks. The density, sonic, neutron and resistivity logs were used as input data while the core data was the output/target. The Well-A data was divided into 3-sets: X% for training, Y% for validation and Z% for testing. The values for X, Y and Z were decided by MATLAB, because the data split was set to *dividerand*. Before training the entire data set was processed using network processing function: *mapminmax*.

Mapminmax is a one of function commonly used in MATLAB for data pre-processing. It performs normalization and linear scaling of data. The mapminmax function helps normalizes the input/target data to a specified range for fast training. The equation for mapminmax function is:

$$y = \frac{(y_{max}-y_{min})*(x-x_{min})}{(x_{max}-x_{min})+y_{min}} \quad (3)$$

3.2 Model Development

3.2.1 MLP Network Architecture

Seven (7) candidate Multiple Layer Perceptron (MLPs) were created and configured as feed forward network in MATLAB Neural Network Toolbox (*nnstart*). Each candidate MLP network had 3-layers: input layer, hidden layer and output layer as represented in Figure 2. The hidden layer consists of 15-30 neurons. Tansig was used as transfer function for neuron in the hidden layer, while purelin is used for output layer.

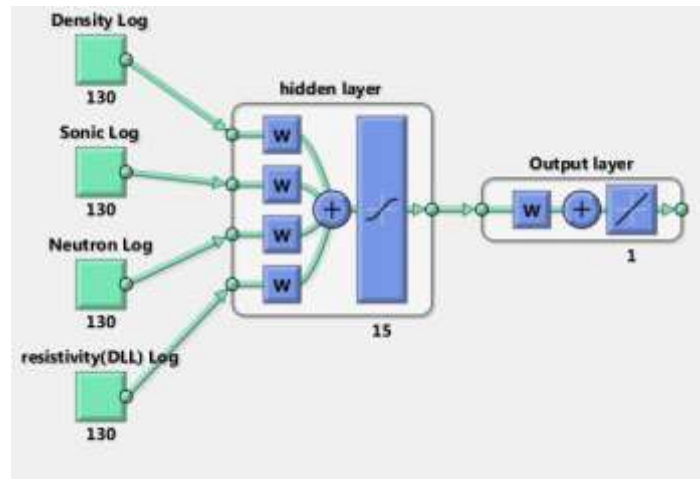


Figure 2: A typical neural network architecture for MLP

3.2.2 Ensemble Network (EN)

All the seven (7) MLPs were randomly initiated and trained multiple times using the Levenberg-Marquardt (*Trainlm*) training function. This training function is the fastest and most suited for pattern recognition task.

In the Toolbox, network performance is evaluated using mean squared error (MSE) during validation and Pearson Correlation Coefficient (R) during testing. The criteria used to determine performance of network is discussed in section 5.2. When the trained networks, four (4) MLPs which had best performance, were selected and joined to form the ensemble network (EN). They are also called the expert networks.

3.2.3 Simulation of EN

The four (4) expert networks which formed the ensemble network (EN) were also randomly initiated and trained multiple times on same training data. The Levenberg-Marquardt training function was also used. Their respective outputs were added and the method of ensemble averaging (equation-2) was used to compute the output of the ensemble network. The EN was also simulated with the test data.

3.2.4 Predicting Water Saturation with EN

The unseen Well-B dataset was deployed in the trained ensemble network to predict water saturation. The water saturation values predicted by the ensemble network, expert network and the Empirical Equation (EE) derived results were compared to that of core data.

4. DISCUSSION AND RESULTS

4.1 Dataset

The limited training and testing dataset used in this research was a major challenge because robust neural network deployment require massive amount of dataset. To solve the problem of over fitting, high variance and local minima that plagues limited data, we used three (3) approaches: (i) Before trainings, the network were randomly initiated, this ensured the network weight were randomly assigned when training begins, so that the network converges to different solutions (ii) each network configuration was trained several times, in order to obtain network with best network performance, thereby reducing the chance of being stuck in a local minima (iii) setting the data split to **dividerand** enabled MATLAB to train and test the network with new data set for each run, thereby reducing the chance of over fitting.

4.2 Evaluating Network Performance

Performance of network was decided by two factors: closeness of predicted output to actual value and the correlation between actual value and predicted value. The **plotperformance** function generates a plot of mean squared error (mse) versus epoch, this helps to visualize closeness between actual value and predicted output. The **plotregression** function creates a regression plot of the network predicted output

versus the actual value. In this study, network output is the predicted value for water saturation. The actual/target value is the water saturation for core. The criteria used to determine network performance is given in table 1.

Table 1: Criteria for Network Performance

Parameter	Poor Network	Good Network	Excellent Network
Correlation (R)	0.00-0.49	0.50-0.79	0.80-1.00
MSE	0.1-0.01	0.001-0.01	0.0001-0.001

4.3 Evaluating performance of candidate MLPs and Expert networks Performance of the seven (7) MLP network was tested using well log data consisting of 50-samples of the unseen well-B data. The density, gamma ray, sonic and resistivity logs were used as input variables and core data as the target output. Table 2 shows the MSE and R for the seven MLP networks ranking. From the table, MLP-4, MLP-5 and MLP-7 display excellent correlation (R) between predicted and actual water saturation values. Also, the MSE values are good, indicating very small error. MLP-3 has small MSE value, but the correlation to core data is good. Since MLP-4, MLP-5, MLP-7 and MLP-3 display acceptable values, these models possess the capacity to effectively map the nonlinear relationship between well log parameters and reservoir water saturation. Therefore, they will be joined to form the ensemble network. Figure 2 shows the R-values between network and target for MLP-1, MLP-2, MLP-3 AND MLP-4, figure 3 show the R-values for MLP-5, MLP-6 AND MLP-7

Table 2: MSE, R and Ranking for the 7-MLP Networks.

Network	MSE	R	Ranking	Remarks
MLP-1	0.01560	0.7194	3	
MLP-2	2.56100	0.7074	1	
MLP-3	0.00872	0.7592	4	4th
MLP-4	0.00021	0.8441	7	1st (Best-Expert)
MLP-5	0.00032	0.8038	5	3rd
MLP-6	0.07510	0.7127	2	
MLP-7	0.00037	0.8064	6	2nd

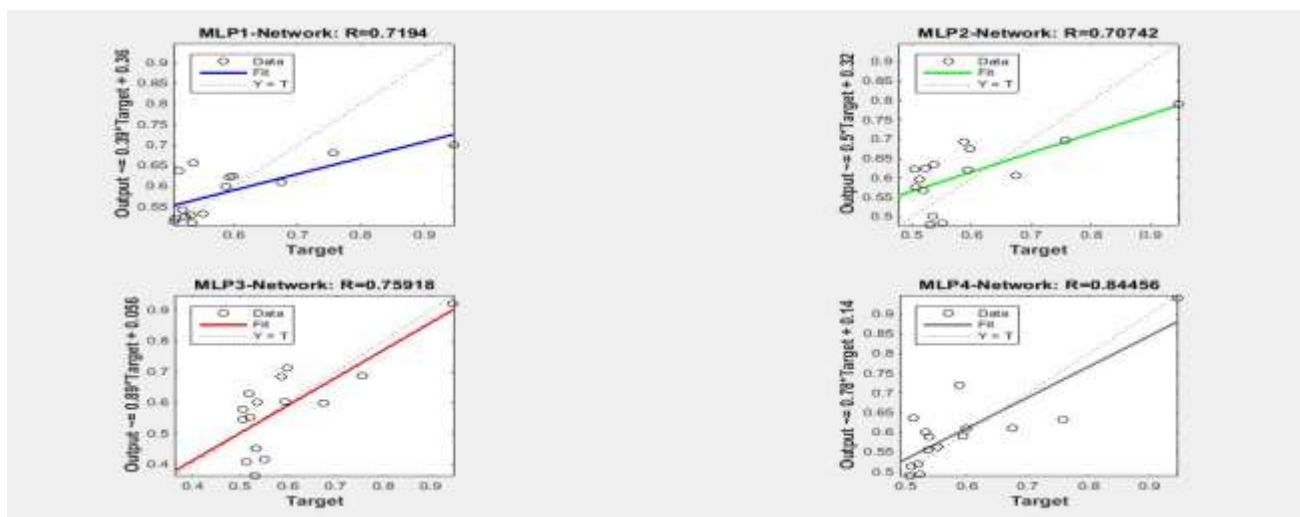
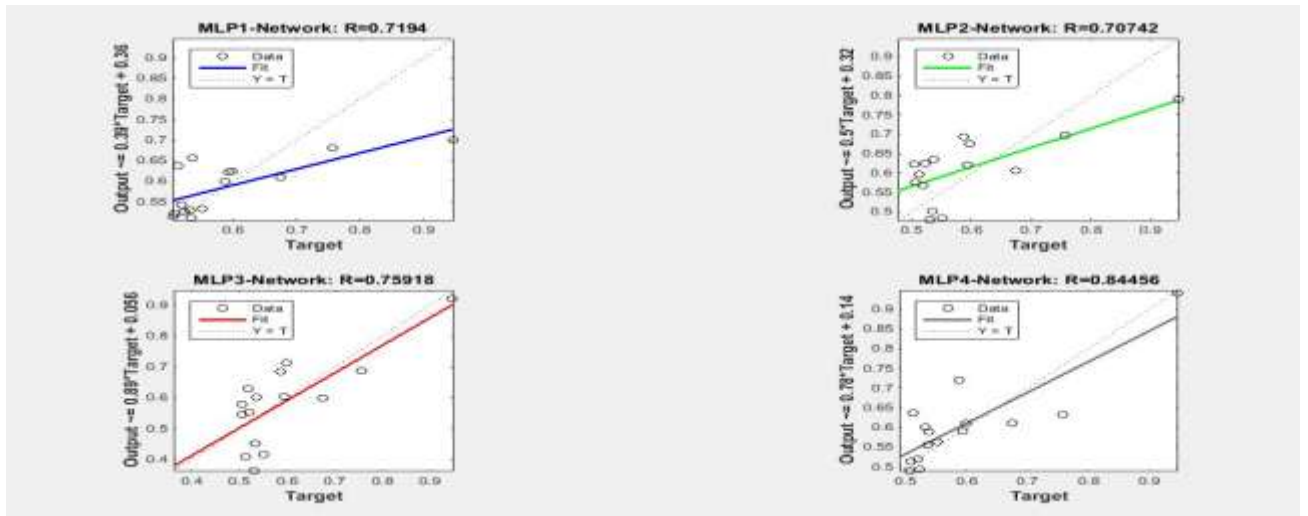


Figure 3: Post regression plots for MLP-1, MLP-2, MLP-3 and MLP-4 Network

Figure 4: Post regression plots for MLP-5, MLP-6, and MLP-7 Network

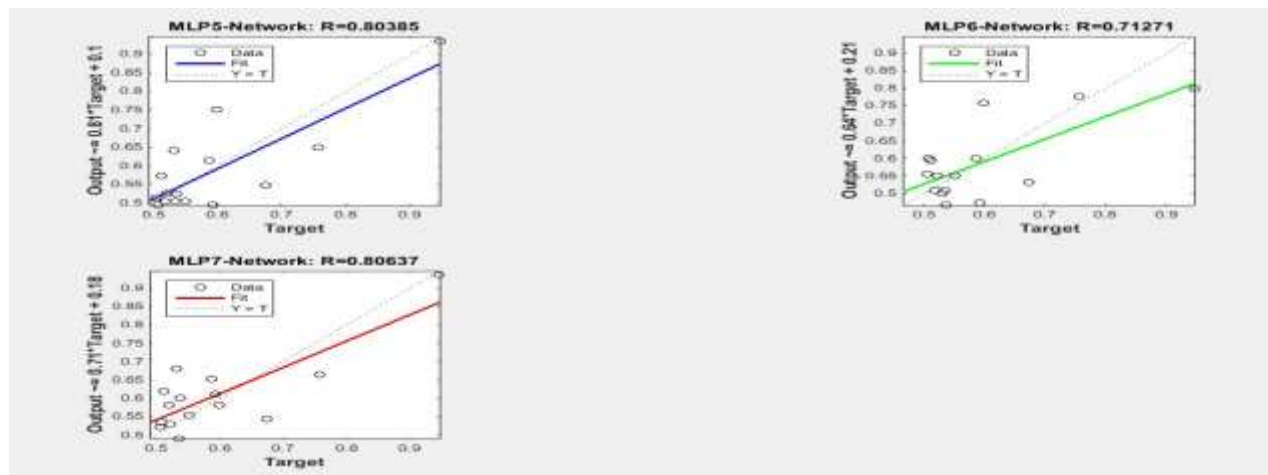


4:4 Evaluating Performance of Expert networks

Performance of the four (4) expert networks was tested using well log data consisting of another 50-samples of the unseen well-B data. Table 3 shows the MSE and R and ranking for the ensemble and expert networks. From the table, the ensemble network has the best correlation with the core data, therefore it possesses the capacity to map the relationship between the well logs and the reservoir water saturation. It marginally out performs MLP-4, which is the best expert network. Figure 5 show the R-values for the ensemble and expert networks

Table 3: MSE, R and ranking for the ensemble and expert Networks.

Network	(MSE)	(R)	Ranking	Remark
Ensemble Network	0.00005	0.9600		
MLP3	0.00075	0.8329	1	
MLP4	0.00034	0.9387	4	Best Expert
MLP5	0.00041	0.9015	2	
MLP7	0.00050	0.9076	3	



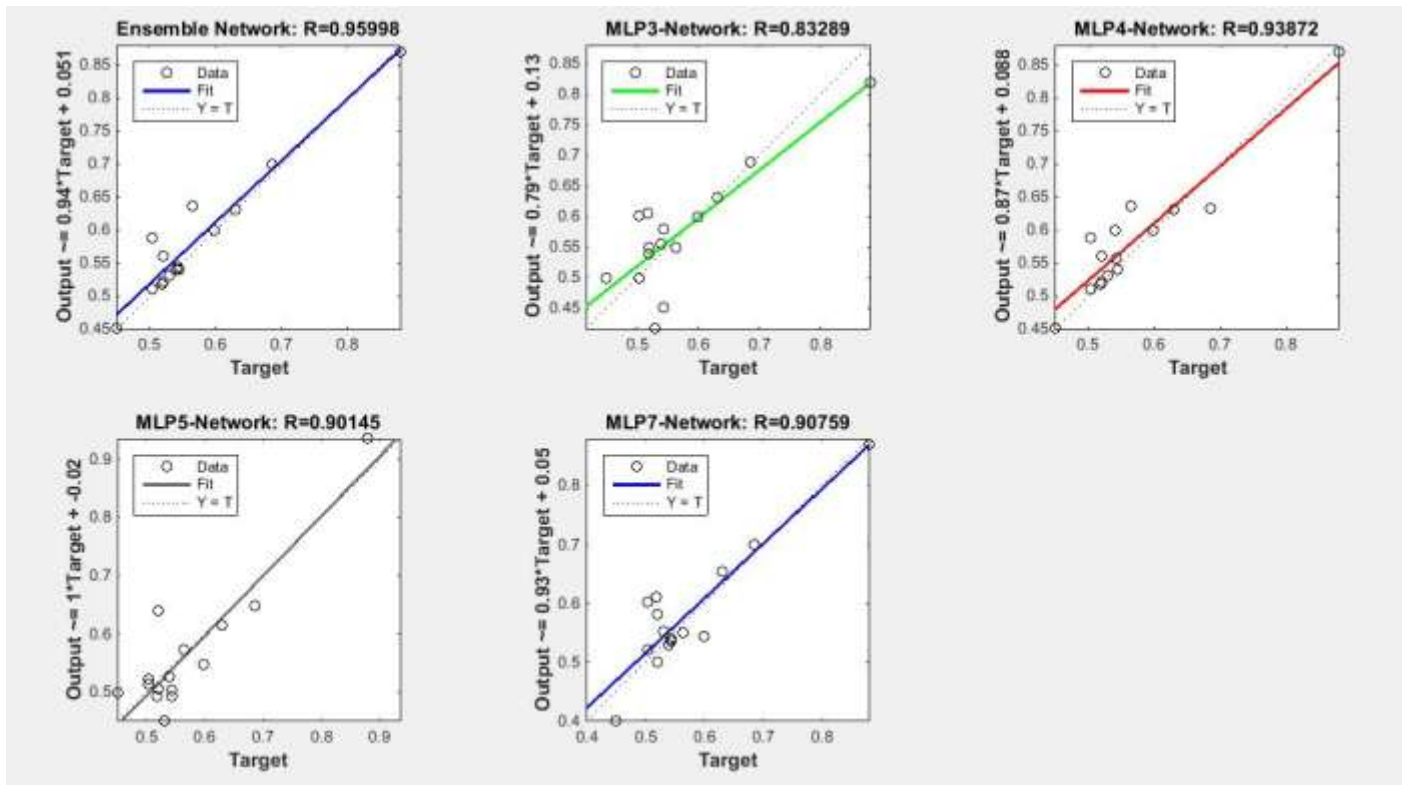


Figure 5: Post regression plots for Ensemble network, MLP-3, MLP-4(best expert), MLP-5 and MLP-7

4:5 Comparing Ensemble Network, Best Expert and EE-Derived Values with Water Saturation of Core

Table 4. Correlation (R) of Ensemble network, Best Expert and EE-derived Values with water saturation of core

Network/Data	(R)	Remark
Core Data	1	
Ensemble Network	0.9796	Best model
Best Expert	0.9557	
(Empirical Equation Derived S_w)	0.9228	

Performance of the ensemble and best expert network was tested using 88 samples of well log data from the unseen well-B data and the results were compared to water saturation values for core data and empirical equation. Table 4 shows the ensemble model displayed the highest correlation (R) to core data. The R^2 values for ensemble network (0.96), best expert network (0.91) and EE-Derived (0.85) shows that the ensemble model fits the core water saturation strongly.

Despite the challenges posed by limited dataset size, the improved performance of the ensemble model over others can be attributed to the diversity among the constituent experts, multiple trainings of expert, random initialization of training and selective combination of the best-performing experts.

Figure 6 shows the Post regression plots for Ensemble network, best expert and EE-derived Values. Plots of Coefficients of Determination (R^2) for Ensemble Network (Figure 7), best expert (Figure 8) and EE-derived water saturation (Figure 9) are also shown.

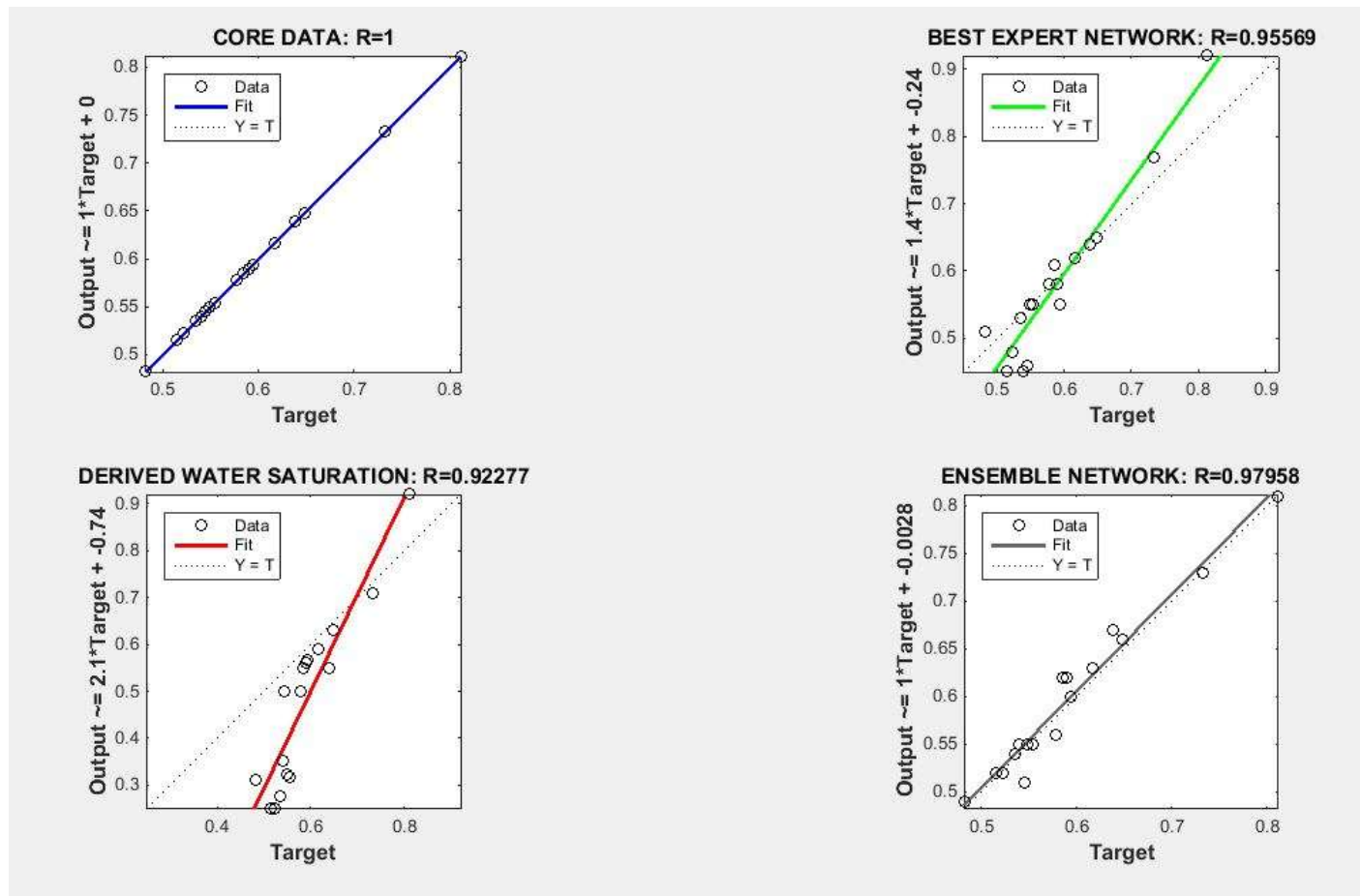


Figure 6: Post regression plots for Ensemble network, best expert and EE-derived water saturation

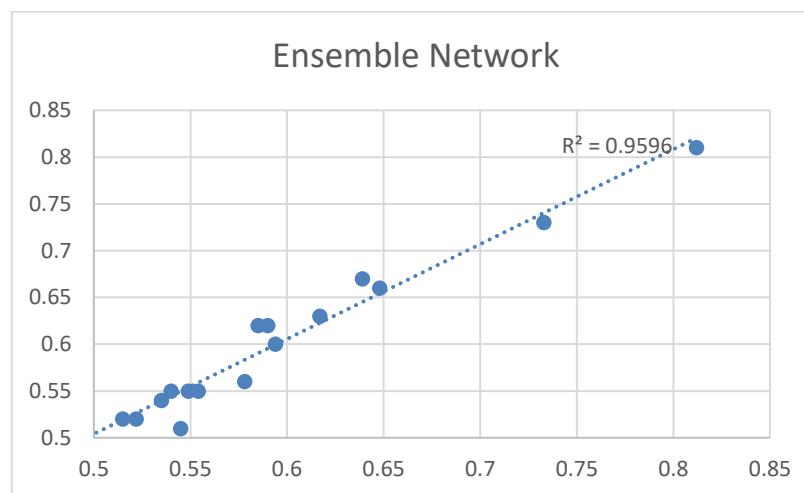


Figure 7: Coefficient of Determination (R^2) for Ensemble Network

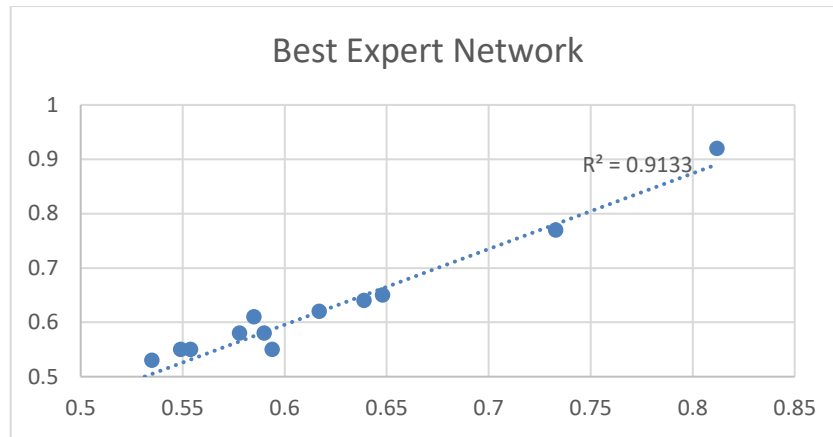


Figure 8: Coefficient of Determination (R^2) for best expert Network

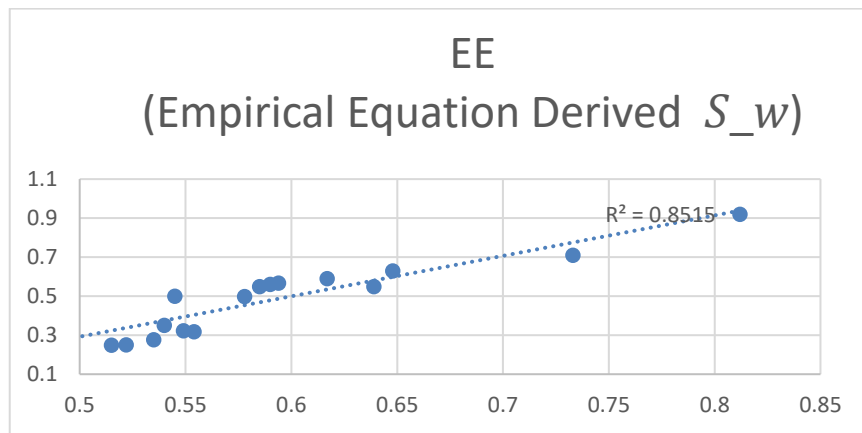


Figure 9: Coefficient of Determination (R^2) for Empirical Equation Derived water saturation

5. CONCLUSION

This study presented an ensemble neural network as a potential data driven tool for predicting water saturation from well logs. The results obtained showed that water saturation values predicted by the ensemble network had the best fit with core data. The deployment of multiple trainings of the neural network, random initialization of training, and random sharing of training data provides a reliable framework for handling small datasets and improving prediction performance.

6. REFERENCE

- [1] Ezekwe, Nnaemeka, (2011) "Petroleum reservoir engineering practice", Pearson Education, Inc.
- [2] Nabil Al-Bulushi, Peter R. King, Martin J. Blunt and Martin Kraaijveld (2009), "Development of Artificial Neural Network Models for Predicting Water Saturation and Fluid Distribution" Journal of Petroleum Science and Engineering, vol 68, 197–208
- [3] Ahmed Farid Ibrahim, Salaheldin Elkatatny and Mustafa Al Ramadan (2022), "Prediction of Water Saturation in Tight Gas Sandstone Formation Using Artificial Intelligence" American Chemical Society, (ACS Omega) 7, 215–222
- [4] Onwusinkwue S., Azi S, Onwusinkwue N, and Ojo K. (2025), "Predicting Permeability Of Reservoir using Committee Neural Network", Transaction of National Association of Mathematical Physics, 22, (125-134)

- [5] Okon, A. N., Adewole, S. E., and Uguma, E. M. (2021). “Artificial neural network model for reservoir petrophysical properties: Porosity, permeability and water saturation prediction” *Model. Earth Syst. Environ.* 7 (4), 2373–2390. <http://doi:10.1007/s40808-020-01012-4>
- [6] Soleimani F., Hosseini E and Hajivand F. (2020). “Estimation of reservoir Porosity using analysis of eismic attributes in an Iranian oil field” *J.Pet.Explor. Prod.Technol.* 10(4), 1289–1316. doi:10.1007/s13202-020-00833-4
- [7] Haykin S. (2009), “Neural networks and learning machines” 3rd edition, Pearson Education, Inc., Upper Saddle River, New Jersey 07458.
- [8] Wood DA (2019), “Predicting Porosity, Permeability and Water Saturation Applying An Optimized Nearest Neighbour, Machine-Learning and Data-Mining Network Of Well-Log Data”. *J Pet Sci Eng* <https://doi.org/10.1016/j.petrol.2019.106587> Engineering 205, 108869
- [9] Rostamian, A., Heidaryan, E., Ostadhassan, M., (2022) “Evaluation of Different Machine Learning Frameworks to Predict CNL-FDC-PEF Logs Via Hyperparameters Optimization and Feature Selection” *J. Petrol. Sci. Eng.* 208, 109463.
- [10] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al Dujaili, Ye Duan, Omran Al, Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana, Amid and Laith Farhan (2021) “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, Big Data “(2021) 8:53 <https://doi.org/10.1186/s40537-021-00444-8>
- [11] Alpna Bhatt (2002) “Reservoir properties from well logs using neural networks” Ph.D Thesis, Department of Petroleum Engineering and Applied Geophysics, Norwegian University of Science and Technology
- [12] Ahmadi, M. A., and Chen, Z., (2019). “Comparison of Machine Learning Methods for Estimating Permeability and Porosity Of Oil Reservoirs Via Petrophysical Logs”. *Petroleum Journal* 5, 271–284. doi:10.1016/j.petlm.2018.06.002
- [13] Sun K and Dong L (2022) “A New Development Algorithm for Permeability Prediction: A New Milestone”. *Front. Ecol. Evol.* 10:1066800. doi: 10.3389/fevo.2022.1066800
- [14] Sambo C, Hermana M, Babasafari A, and Janjuhah H. T (2018), “Application of Artificial Intelligent Methods for Predicting Water Saturation From New Seismic Attributes”. In: Paper SPE OTC-28221 presented at the offshore technology conference Asia, Kuala Lumpur, Malaysia, 20–23 Mar. <https://doi.org/10.2118/28221-MS>
- [15] Wood DA (2019), “Predicting porosity, permeability and water saturation applying an optimized Nearest neighbour, machine-learning and data-mining network of well-log data. *J Pet Sci Eng* 184:1–17. <https://doi.org/10.1016/j.petrol.2019.106587> Engineering 205, 108869