PREDICTION OF POROSITY AND PERMEABILITY OF OIL AND GAS RESERVOIRS USING SUPPORT VECTOR MACHINES AND ARTIFICIAL NEURAL NETWORKS: A COMPARATIVE STUDY

Ewenla A.*, Anifowose F.^, Akanbi L.*+ , Oluwatope O.A.*+, Aderounmu G.A.*+

**Department of Computer Science & Engineering, Obafemi Awolowo University, Nigeria*
^Center for Petroleum & Minerals, King Fahd University of Petroleum and Minerals, Saudi Arabia.
**honourablevc@yahoo.com, ^anifowo@kfupm.edu.sa, +[lakanbi,aoluwato,gaderoun@oauife.edu.ng]*

Abstract. There has been a persistent quest for improved prediction of the properties of oil and gas reservoir. The performance of existing techniques has been acceptable but needs to be improved upon. Petroleum reservoir modeling has evolved from the use of empirical and statistical tools to the embrace of computational intelligence techniques especially Artificial Neural Networks (ANN). ANN has been used extensively in petroleum engineering applications but has a lot of limitations. This paper presents a comparative study of the application of ANN and Support Vector Machines (SVM) in the prediction of porosity and permeability of petroleum reservoirs. SVM has been reported in literature to be efficient, easy to train and resistant to overfitting. The petroleum industry has not adequately benefitted from the excellent generalization capability of SVM. Datasets from different petroleum reservoirs were used in this study to develop, train and evaluate the comparative performance of ANN and SVM models. The results showed that the SVM models performed better, in terms of higher correlation coefficients, lower root mean squared errors and less execution time. Hence, we present SVM as a possible alternative to ANN, especially, in the prediction of petroleum reservoir properties.

Keywords: Artificial Neural Networks, Support Vector Machines, reservoir characterization, porosity, permeability.
PACS: 07.05.Tp, 89.20.Ff, 07.05.Mh, 89.30.aj, 91.60.Np

INTRODUCTION

Porosity and permeability are two important properties of petroleum reservoirs that relate to the amount of fluid in a reservoir and its ability to flow. These properties have significant impact on petroleum field operations and reservoir management. They both serve as standard indicators of reservoir quality in the oil and gas industry.

Porosity is the percentage of voids and open spaces in a rock or sedimentary deposit. The greater the porosity of a rock, the greater its ability to hold water and other materials, such as oil. It is an important consideration when attempting to evaluate the potential volume of hydrocarbons contained in a reservoir. Permeability is the ease with which fluid is transmitted through a rock's pore space. It is a measure of how interconnected the individual pore spaces are in a rock or sediment. It is a key parameter associated with the characterization of any hydrocarbon reservoir. In fact, many Petroleum Engineering problems cannot be solved without having an accurate permeability value [1].

Until the embrace of Computation Intelligence (CI) techniques in the petroleum industry, modeling of petroleum reservoirs has been done with empirical equations and later with statistical tools. It has been argued that such natural phenomena as porosity and permeability cannot be adequately estimated by linear relations [1, 2]. The CI techniques achieved this by establishing non-linear relations between the log measurements and the core values for prediction [3-7]. However, Artificial Neural Networks (ANN) has shown better prediction accuracies. ANN has been successfully applied on petroleum reservoir characterization tasks over the years. Despite this, it has been observed to have some limitations [8]. Support Vector Machines (SVM) was introduced with features that overcame some of the limitations of ANN. SVMs have performed well in a good number of applications such as reported in [1, 2].

This paper focuses on the comparative performance of ANN and SVM models using datasets that are representative of most geological formations in the world. To achieve this aim, Section 2 presents a survey of ANN and SVM. Section 3 describes the research methodology, the datasets and the criteria to evaluate the models. Section 4 presents the results of the study with a detailed discussion while conclusion is presented in Section 5.

LITERATURE SURVEY

This section reviews the literature about the architecture and existing applications of ANN and SVM within and outside the petroleum industry.

Theory of Artificial Neural Networks

ANN is made up of several layers of neurons (nodes) interconnected by links using weights. Multi-Layer Feed-Forward Neural Network (MLFFNN) is the most popular and widely used paradigm. It contains one input layer, one output layer and one or more intermediate hidden layers. Signals pass through the network in the forward direction, starting from the input layer, moving to the hidden layer (or layers) and then to the output layer. In the hidden and output layers, each neuron sums the weighted values (signals) from the previous layer and puts the result through a transfer function, which is also called activation function, to give a scalar output value (signal) [8]. Figure 1 shows a simplified structure of an Artificial Neural Network, which can be mathematically generalized as:

$$y_i = f\left(\sum_k w_{ik} x_k + \mu_i\right) \tag{1}$$

where x_k are inputs to the neurons, w_{ik} are weights attached to the inputs to the neuron, μ_i is a threshold, offset or bias, $f(\bullet)$ is a transfer function and y_i is the output of the neuron. The transfer function $f(\bullet)$ can be any of: linear, non-linear, piece-wise linear, sigmoidal, tangent hyperbolic and polynomial functions.

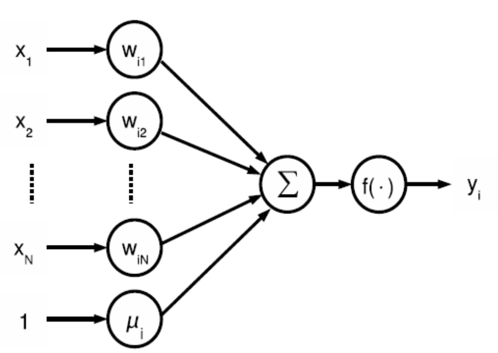

FIGURE 1. Architecture of Artificial Neural Networks [8].

Existing Work on Artificial Neural Networks

A good number of efforts have been made in the application of ANN in various fields and especially in oil and gas reservoir characterization. Ref. [3] used ANN as a multivariate correlative tool to predict permeability from petrographic data using Fuzzy Logic to screen and rank the predictor variables with respect to the target variable. The result demonstrated the generalizing capability of ANN. Ref. [4] introduced a new Neural-Fuzzy technique combined with Genetic Algorithms in the prediction of permeability in petroleum reservoirs. The methodology involved the use of ANN to generate membership functions and to approximate permeability automatically from well logs. The results showed that the integrated tool gave the smallest error on the unseen data when compared to similar algorithms.

Ref. [5] used ANN and Fuzzy Logic to characterize naturally fractured reservoirs. The results showed that the proposed approach is a practical methodology to map the fracture network. In a similar study, [6] suggested an intelligent technique using Fuzzy Logic and ANN to determine reservoir properties from well logs. The results showed that the technique estimated the reservoir properties more accurately and reliably than conventional computing methods.

Ref. [7] proposed a new method for the auto-design of ANN based on Genetic Algorithm. The design of the topology and parameters of the ANN model was done by using Genetic Algorithms, in order to improve the effectiveness of forecasting when ANN was applied to a permeability predicting problem from well logs. Though, the GA-ANN performed better but required more time and memory since GA is known as a memory-intensive and time-consuming optimization tool.

Despite the applications of ANN, it was reported to suffer from the following deficiencies [8]:

- There is no general framework to design the appropriate network for a specific task.
- The number of hidden layers and hidden neurons of the network architecture are determined by trial and error.
- A large number of parameters are frequently required to fit a good network structure.
- ANN uses pre-defined activation functions without considering the properties of the phenomena being modeled.

Various studies to address these problems through the development of other algorithms such as Cascade Correlation and Radial Basis Function did not improve its overall performance [8 - 10].

Theory of Support Vector Machine

SVM is a set of related supervised learning methods used for classification and regression. They belong to a family of generalized Linear Classifiers. They can also be considered as a special case of Tikhonov Regularization. SVMs map input vectors to a higher dimensional space where a maximal separating hyperplane is constructed. The generalization ability of SVMs is ensured by the special properties of the optimal hyperplane that maximizes the distance to training examples in a high dimensional feature space. Recently, a new ε-sensitive loss function technique that is based on statistical learning theory, and which adheres to the principle of structural risk minimization, seeking to minimize an upper bound of the generalization error was developed. This gave rise to the new technique called Support Vector Regression (SVR). It has been shown to exhibit excellent performance [11]. A conceptual framework of SVM is shown in Figure 2.

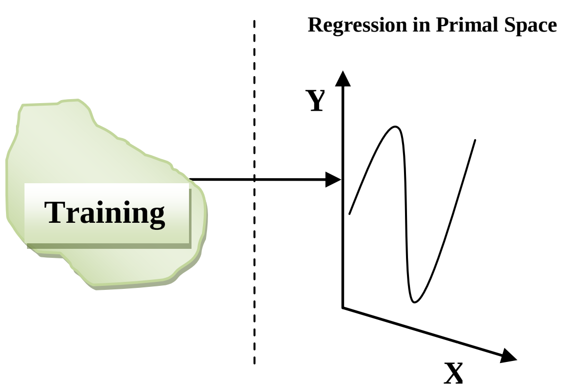

FIGURE 2. Mapping Input Vectors to a Higher Dimensional Space in SVM [11].

Existing Work on Support Vector Machines

Due to the above merits, SVM has been successfully applied in many areas such as decision support, software reliability identification, pattern recognition, and in the prediction of oil and gas properties.

Ref. [1, 2, 11] demonstrated the special features of SVM in terms of its ability to handle small dataset using the scale proposed by [12]. In terms of both execution time and correlation coefficient, they presented SVM to be a very close match to Functional Networks with very high mutually competitive edges. In a related study, [13] proposed ANN, SVM and Functional Networks to predict the Pressure-Volume-Temperature (PVT) properties of crude oil. The result showed that SVR and FN are competitive but SVM has the overall best result for gas and oil prediction. Ref. [14] proposed the application of SVM for prediction of toxic activity with different datasets. When compared with ANN and RBF models, SVM gave the highest correlation coefficient. Other successful applications of SVM include [15, 16] .

RESEARCH METHODOLOGY

This section describes the setup of the experiment. First, the datasets are described. Then, the criteria used to evaluate the performance of the models are discussed. Lastly, the succinct account of the experiment was presented.

Description of Datasets

For the purpose of evaluating the performance of these techniques and for effective comparison with other possible existing work, the same datasets used by [1, 2] were used. The datasets consist of two sets of well logs obtained from a drilling site containing three wells for porosity in the Northern Marion Platform in North America (Site 1) and another set of well logs from a drilling site containing three wells for permeability in the Middle East (Site 2). The datasets from Site 1 contain six predictor variables for porosity while the datasets from Site 2 contain eight predictor variables for permeability. These are shown in Table 1 and 2.

TABLE 1. Predictor Variables for Porosity (Site 1).

1.	Core
2.	Top Interval
3.	Grain Density
4.	Grain Volume
5.	Length
6.	Diameter

TABLE 2. Predictor Variables for Permeability (Site 2).

1.	Gamma Ray Log
2.	Porosity Log
3.	Density Log
4.	Water Saturation
5.	Deep Resistivity
6.	Microspherically Focused Log
7.	Neutron Porosity Log
8.	Caliper Log

The datasets were divided into training and testing subsets using a Stratified Sampling technique [1, 2, 11, 12, 17 - 19]. Using this technique, a randomized sampling of 70% of the entire data was selected for training and the remaining 30% left for testing. The number of data points and their corresponding stratifications are shown in Table 3.

TABLE 3. Division of Datasets into Training and Testing Subsets.

	Porosity			**Permeability**		
Wells	**1**	**2**	**3**	**1**	**2**	**3**
Data Points	415	285	23	355	477	387
Training (70%)	291	200	16	249	334	271
Testing (30%)	124	85	7	106	143	116

Criteria for Performance Evaluation

The performance of the models was evaluated using the correlation coefficient (CC), root mean-squared error (RMSE) and execution time (ET). CC measures the statistical correlation between the predicted and actual values. RMSE is one of the most commonly used error measures of success for numeric prediction. It computes the average of the squared differences between each predicted value and its corresponding actual value. ET is simply the total time taken for a technique to run from the beginning to its end.

Models Implementation Strategy

The architecture of the ANN models used in this study consists of six neurons in the input node for porosity and eight neurons for permeability respectively. The inputs are the reservoir parameters for the respective oil and gas reservoir sites. The middle layer consists of the hidden layer with neurons that are trained to generate a mapping between the input values from the input layer and the predicted porosity and permeability values in the output layer. The optimal configuration used was 6-40-40-1 for porosity and 8-40-40-1 for permeability. The list below summarizes the optimized parameters:

- Type = Feed-Forward Back-Propagation
- Number of Hidden Layers = 2
- Number of Neurons in each Hidden layer = 40
- Training Algorithm = Levenberg-Marquardt (Trainlm)
- Transfer Functions: 'tansig' for the hidden layer and 'purelin' for the output layer.
- Training error goal: 1e-5

For the SVM model, the least-square fitting algorithm, which attempts to minimize the sum of squared errors (SSEs) of training samples while simultaneously minimizing the margin error, was used. Extensive empirical comparisons had shown that LS-SVM obtains good performance on various classification and regression problems [16]. The kernel used was polynomial, which is of the form:

$$k\,(x, y) = (x^T y + 1)^d \qquad (2)$$

and the error, designated by ε, was set to 0.2. Other optimized parameters of the SVM model include:

- C = 450;
- lambda = 1e-7;
- epsilon = 0.2;
- kerneloption = 0.30;
- kernel = 'poly';
- verbose = 1;

RESULTS AND DISCUSSIONS

After running the experiment as described in the previous section separately for porosity and permeability, a comparative performance analysis was carried out and the results are displayed in Figure 3 through 8. Since the page limitation here will not accommodate the training results, we present the comparative analysis of the generalization capability of the models. The target of any simulation and modeling task is to see how the model generalizes on new cases.

For porosity, the CC comparative results showed that in terms of training, ANN performed better than SVM in Wells 1 and 2 where the datasets are large. However, in testing, SVM performed better in all 3 Wells (Figure 3). The performance of SVM in the Well 3 with small dataset shows the special capability of SVM to handle small datasets better than ANN. It is thus expected that the RMSE for ANN will be lower for training than those of SVM except for Well 3 where the dataset is small. However, as shown in Figure 4, the RMSE for the SVM model is lower than those of ANN. In terms of execution time, Figure 5 showed that SVM executed faster than ANN in addition to its ease of training and excellent performance.

For permeability, the same trend observed with porosity continued. The CC comparative results showed that SVM performed better in all 3 Wells (Figure 6) with higher correlation coefficient. Similarly, the RMSE for ANN was higher than those of SVM, making the latter the better model (Figure 7). In terms of execution time, Figure 8 showed that SVM still executed faster than ANN.

Presented in terms of percentages, the SVM model was 7% more accurate than the ANN model in terms of correlation coefficient on the Porosity Well 1 and 2 datasets. For the Porosity Well 3 dataset, it was 48%. Almost the same magnitude applies to the RMSE criterion for the Porosity Well 1 and 2 datasets. For the Porosity Well 3 dataset, SVM has about 85% reduction in error than the ANN model. In terms of execution time, SVM models were faster than the ANN models in hundreds of times.

For permeability, the SVM models had about 13.5% improvement in predictive performance than the ANN models in terms of correlation coefficient. Trends similar to the porosity datasets were also observed with the permeability datasets in terms of RMSE and execution time.

The excellent performance of the SVM models is due to its confirmed robustness and especially its superior capability to handle the small Porosity Well 3 dataset. The overall comparative result perfectly agrees with literature that ANN needs more data to obtain a good fit.

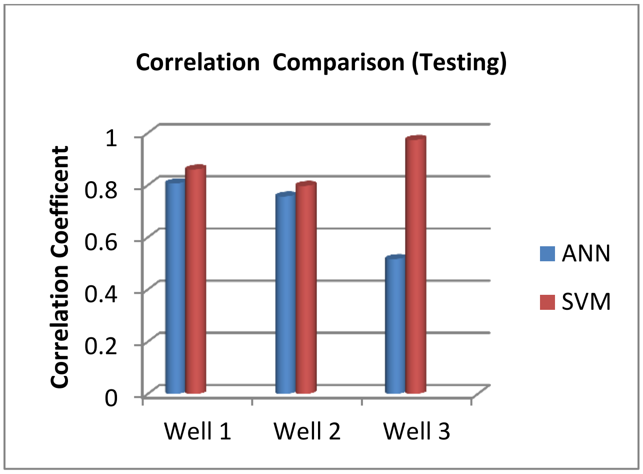

FIGURE 3. Comparative Plot of Porosity Testing CC for all Wells.

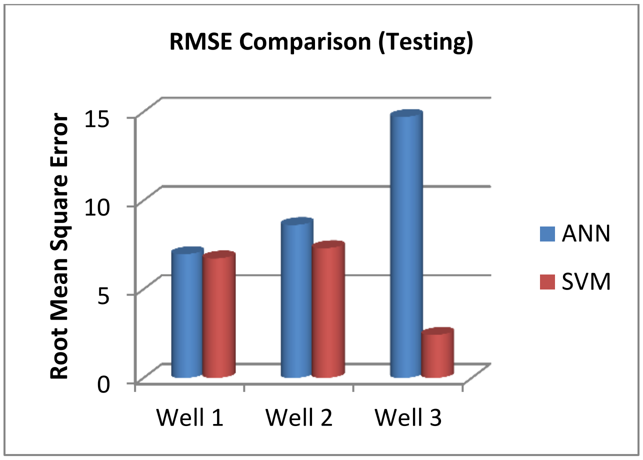

FIGURE 4. Comparative Plot of Porosity Testing RMSE for all Wells.

FIGURE 5. Comparative Plot of Porosity Testing ET for all Wells.

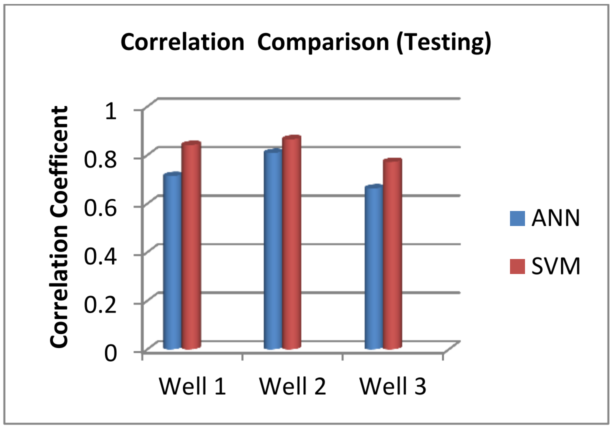

FIGURE 6. Comparative Plot of Permeability Testing CC for all Wells.

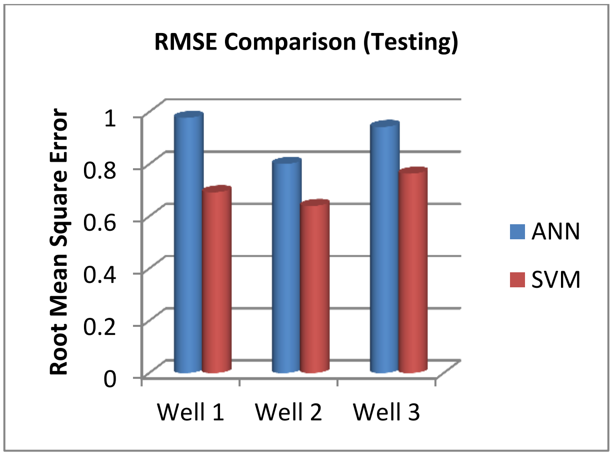

FIGURE 7. Comparative Plot of Permeability Testing RMSE for all Wells.

FIGURE 8. Comparative Plot of Permeability Testing ET for all Wells.

Conclusion

This study has presented SVM as a desirable alternative to ANN. Despite the efforts to handle the limitations of ANN, SVM has proven to be a more robust and versatile tool. The better predictive performance of the generalization capability SVM can be attributed to its ability to handle small datasets and the inability of ANN to converge on global optima in addition to requiring much data samples to obtain a good fit.

The higher speed of execution of SVM can equally be attributed to the complexity of the learning algorithm of ANN which results in its taking more time to obtain a good fit than the simpler architecture of SVM.

The superior features of SVM such as ease of training, simple architecture and stability in its convergence to global optima have made it a more desirable and in fact, a better alternative to ANN, especially in the field of oil and gas exploration where higher accuracy predictions are needed for continued production of the oil the world needs.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support provided by:

- Obafemi Awolowo University Ile-Ife, Nigeria, through the step B project grant from the Federal Ministry of Education. For the establishment of the Center of Excellence in Software Engineering.
- King Abdulaziz City for Science and Technology (KACST) through the Science &

Technology Unit at King Fahd University of Petroleum & Minerals (KFUPM) for funding this work under Project No. 08-OIL82-4 as part of the National Science, Technology and Innovation Plan.

REFERENCES

1. F. Anifowose and A. Abdulraheem, "Prediction of Porosity and Permeability of Oil and Gas Reservoirs using Hybrid Computational Intelligence Models" in *Proceeding of SPE North Africa Technical Conference and Exhibition*, Cairo, Egypt, 14-17 February 2010.
2. A. Maqsood and C. Adwait, "Using Artificial Intelligence to Predict Permeability from Petrographic Data", *Computers & Geosciences* **26**, 915-925 (2000).
3. T. Helmy, F. Anifowose and K. Faisal, "Hybrid Computational Models for the Characterization of Oil and Gas Reservoirs", *Expert Systems with Applications* **37**, 5353-5363 (2010).
4. H. Yuantu, G. Tom and W. Patrick, "An Integrated Neural-Fuzzy-Genetic-Algorithm using Hyper-Surface Membership Functions to Predict Permeability in Petroleum Reservoirs", *Engineering Applications of Artificial Intelligence* **14**, pp. 15-21 (2001).
5. E. Abdelkader, T. Djebbar and M. Amine, "Application of Artificial Intelligence to Characterize Naturally Fractured Zones in Hassi Messaoud Oil Field, Algeria", *Journal of Petroleum Science and Engineering* **49,** 122-141 (2005).
6. L. Jong-Se, "Reservoir Properties Determination using Fuzzy Logic and Neural Networks from Well Data in Offshore Korea", *Journal of Petroleum Science and Engineering* **49**, 182-192 (2005).
7. S. Mohsen, A. Morteza, and Y.V. Ali, "Design of Neural Networks using Genetic Algorithm for the Permeability Estimation of The Reservoir", *Journal of Petroleum Science and Engineering* **59**, 97-105 (2007).
8. J.B. Petrus, F. Thuijsman and A.J. Weijters, Artificial Neural Networks: *An Introduction to ANN Theory and Practice*, Springer, 1995, pp. 37-57.
9. M. Bruen and J. Yang, "Functional Networks in Real-Time Flood Forecasting: A Novel Application", *Advances in Water Resources* **28**, 899-909 (2005).
20. E. A. El-Sebakhy, A. S. Hadi and A. K. Faisal, "Iterative Least Squares Functional Networks Classifier", I*EEE Transactions Neural Networks* **18**, 3, 844-850 (2007).
31. F. Anifowose, *Hybrid AI Models for the Characterization of Oil and Gas Reservoirs: Concept, Design and Implementation*, VDM Verlag, Germany, 2009.
42. F. Anifowose and A. Abdulraheem, "How Small is a "Small" Data?" in *Proceedings of the 2nd Saudi Conference on Oil and Gas Exploration and Production Technologies (OGEP 2010)*, Dhahran, Saudi Arabia, 18-20 December, 2010.
53. M. A. Oloso, A. Khoukhi, A. Abdulraheem and M. Elshafei, "Prediction of Crude Oil Viscosity and Gas/Oil Ratio Curves Using Recent Advances to Neural Networks" in *2009 SPE/EAGE Reservoir Characterization and Simulation Conference*, Abu Dhabi, UAE, 19–21 October, 2010.
64. C.Y. Zhao, H.X. Zhang, M.C. Liu, Z.D. Hua and B.T. Fan, "Application of Support Vector Machine (SVM) for Prediction Toxic Activity of Different Data Sets", *Toxicology* **217**, pp. 105-119 (2006).
75. K. Wong, Y. Ong, T. D. Gedeon and C. C. Fung, "Reservoir Characterization Using Support Vector Machines" in *Joint International Conference on Computational Intelligence for Modeling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce*, 2005.
86. P. Xinjun and W. Yifei, "A Normal Least Squares Support Vector Machine (NLS-SVM) and its Learning Algorithm", *Neurocomputing* **72**, 3734-3741 (2009).
97. T. Helmy and F. Anifowose, "Hybrid Computational Intelligence Models for Porosity and Permeability Prediction of Petroleum Reservoirs", *International Journal of Computational Intelligence and Applications* **9**, 4, 313–337 (2010).
108. F. Anifowose and A. Abdulraheem, "A Functional Networks-Type-2 Fuzzy Logic Hybrid Model for the Prediction of Porosity and Permeability of Oil and Gas Reservoirs" in *Proceeding 2nd International Conference on Computational Intelligence, Modeling and Simulation (CIMSim 2010)*, IEEEXplore, Bali, Indonesia, September 28-30, 2010, pp. 193-198.
119. F. Anifowose and A. Abdulraheem, "A Fusion of Functional Networks and Type-2 Fuzzy Logic for the Characterization of Oil and Gas Reservoirs" in *Proceeding the 2010 International Conference on Electronics and Information Engineering (ICEIE 2010)*, Kyoto, Japan, August 1-3, IEEEXplore Vol. 2, 2010, pp. 349-353.