AIP Conference Proceedings

scitation.org/journal/apc

Volume 2563

The 2nd Universitas Lampung International Conference on Science, Technology, and Environment (ULICoSTE) 2021

Bandar Lampung, Indonesia • 27–28 August 2021

Editors • Lusmeilia Afriani, Rudy, Ryzal Perdana, Gede Eka Putrawan and Trio Yuda Septiawan

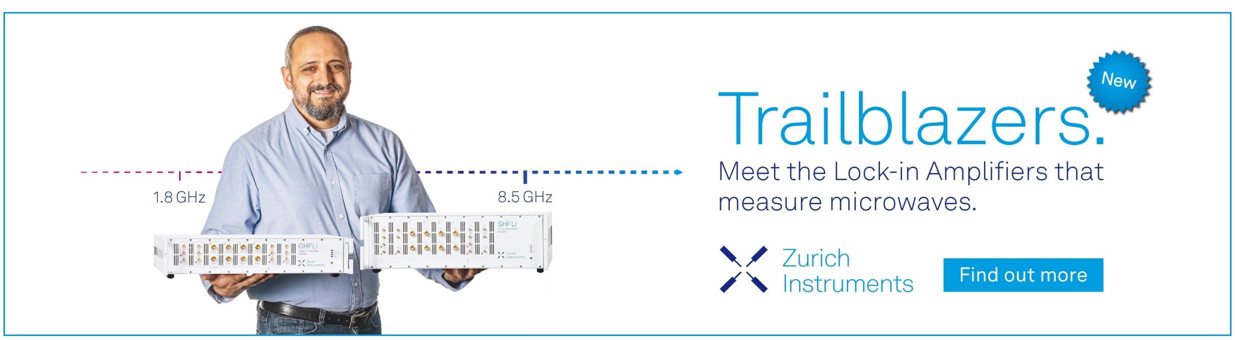
2021
2nd ULICoSTE
UNIVERSITAS LAMPUNG INTERNATIONAL CONFERENCE ON SCIENCE, TECHNOLOGY AND ENVIRONMENT

AIP Publishing

Total organic carbon (TOC) prediction using machine learning methods based on well logs data

Cite as: AIP Conference Proceedings **2563**, 030005 (2022); https://doi.org/10.1063/5.0103209
Published Online: 31 October 2022

Rahmat Catur Wibowo, Ordas Dewanto and Muh. Sarkowi



Total Organic Carbon (TOC) Prediction Using Machine Learning Methods Based on Well Logs Data

Rahmat Catur Wibowo[1, a)], Ordas Dewanto[1, b)], and Muh. Sarkowi[1, c)]

[1]*Geophysical Engineering, Engineering Faculty, Universitas Lampung*
Sumantri Brojonegoro Street, No. 1, Rajabasa, Bandar Lampung City, Lampung, 35145, Indonesia

[a)] *Corresponding author: rahmat.caturwibowo@eng.unila.ac.id*
[b)] *ordasdewanto@gmail.com*
[c)] *muh.sarkowi@eng.unila.ac.id*

Abstract. Evaluation of a source rock can use several parameters, one of which is the determination of Total Organic Carbon (TOC). The determination of TOC is a method that relies on expensive laboratory testing and is limited by the availability of rock samples. TOC prediction using well log data can be performed on most oil and gas wells, which can provide information regarding organic content and continuous data recording. So, the prediction method using well log data is an ideal method to determine TOC in source rock units. The purpose of this study is to predict the TOC value using a well log by applying the machine learning method with the Multi-Layer Perceptron Artificial Neural Network (ANN) technique. Eighteen data samples from the Talang Akar Formation were used for training and testing the MLP-ANN model. The well log data used to predict TOC are density log (RHOB), transit time (DT), deep resistivity (ILD), gamma-rays (GR), and neutron porosity (NPHI), and produce a high correlation (R^2 0.87 and the mean absolute percentage error (AAPE) 10%) against the resulting MLP-ANN model. The TOC prediction technique carried out will help a geophysicist (geophysicist and reservoir geology) to evaluate the source rock in an oil and gas field without the need to have a large number of source rock sample data.

INTRODUCTION

Shale gas is an alternative resource to conventional natural gas and has become a popular resource in oil exploration and production, especially in recent years. This type of resource makes the evaluation of source rock very important to do. Especially in shale, evaluation related to the quality and potential of hydrocarbons is very important and the Total Organic Carbon (TOC) content is the most important parameter. In general, the TOC values contained in the source rock are measured directly in the laboratory, but this method is very limited because it depends on the number of core samples and the cost of analysis is very expensive. The source rock has a specific response to rock density (RHOB), transit time (DT), gamma-rays (GR), and resistivity values (ILD/LLD), which makes it useful for distinguishing surrounding rocks [1-4]. So that the well log data can be used for the evaluation of the source rock. TOC prediction using well logs can provide continuous evaluation of the source rock or target formation in the borehole.

Previous researchers such as Fertl and Rieke (1980) [5] and Fertl and Chilingar (1988) [6] have initiated a TOC evaluation technique using well log data, namely utilizing GR spectral logs in identifying organic-rich rocks and looking for relationships between total GR ratios, U, and Th-K. Schmoker and Hester (1983) [7] introduced the log RHOB approach in estimating TOC, and it is often used in shales. Mendelzon and Toksoz (1985) [8] have another point of view, namely by using a multivariate method to find a quantitative relationship between well log data and TOC values based on core samples which produce a regression equation with a high correlation coefficient. Carpentier et al. (1991) [9] had the idea of obtaining an estimate of the TOC content in situ based on the transit time of the sonic log (DT) and resistivity using the CARBOLOG method. The method explains that the shale layer rich in organic material has the characteristics of a long transit value and a high resistivity value. However, it requires TOC (core)

The 2nd Universitas Lampung International Conference on Science, Technology, and Environment (ULICoSTE) 2021
AIP Conf. Proc. 2563, 030005-1–030005-8; https://doi.org/10.1063/5.0103209
Published by AIP Publishing. 978-0-7354-4237-5/$30.00

data for calibration [10]. In addition, the DlogR technique was introduced by Passey et al. (1990) [11] which performed an overlay between the porosity log and the resistivity log to predict the TOC value qualitatively and quantitatively. Kamali and Mirshady (2004) [12] have a new idea in determining the TOC value in shale gas by combining the DlogR method and the Neuro-Fuzzy approach [13]. Passey et al. (2010) [14] then updated their findings by including organic-rich rocks that have a high maturity level as a calibration of the DlogR method and identify source rocks that have a good maturity level in shale gas. Integration between DlogR and neural networks in predicting TOC values was carried out by Bakhtiar et al. (2011) [4] to evaluate the source rock. So the DlogR technique is a well log-based technique that is most often used in evaluating TOC values. However, the DlogR technique has problems such as having to choose a baseline manually, the TOC value is quite difficult to determine and regionally has variations. Artificial intelligence systems and neural networks have carried out the prediction of TOC, especially in recent years. Kadkhodaie-Ilkhchi et al. (2009) [15] applied machine learning to predict TOC values from well log data or rock physics. An artificial intelligence method based on well log data was also applied by Khoshnoodkia et al. (2011) [16] to determine the TOC value of the Gadvan formation. Some things that must be considered in using neural networks in TOC calculations are that the calculations involve many parameters and are very complex, so choosing the relevant parameters is very important and generally complicated.

TOC predictions in most cases were obtained by reconstructing the multiregression method. In recent years, several direct TOC estimation techniques have been introduced using well logs (geochemical logs). This technique can display TOC values directly without complex calculations [17-18] but is only found in certain wells.

Machine learning (ML) and artificial intelligence (AI) are techniques that combine computing and human intelligence to obtain precise solutions to complex and nonlinear problems [19]. Several journals in the last two decades have published many articles related to AI and ML for regression correlation and multivariable classification. The development of computational techniques in the field of reservoir characterization, utilizing AI-based correlation techniques [20-21], as well as in engineering, and geomechanics. This correlation is also applied to the field of petroleum geochemistry which can be seen in the work of Rahaman and Vasant (2020) [22]. In recent years, a multi-layer perceptron (MLP) artificial neural network (ANN) has been used to predict porosity values by researchers [23]; researchers have also used MLP-ANN to predict permeability [24].

Based on the previous explanation, the prediction of TOC values in shale rich in organic material can be done using well log data. However, the model obtained does not yet have high accuracy unless it has a lot of core data in predicting TOC. Thus, this study aims to predict TOC using well logs in Talang Akar organic material-rich shales using the MLP-ANN technique. This research will utilize ML and evolutionary algorithms to obtain an optimal model.

METHODS

Several conventional well logs such as RHOB, DT, ILD, GR, and NPHI, were used to predict the TOC value, especially in the Talang Akar shale, which had previously been trained on the MLP-ANN model. Some of these well logs were used to estimate TOC (MLP-ANN model training) because of their proven high accuracy (R^2 more than 0.6), especially in evaluating source rock. The target zone has 18 core data samples and log data. The well log data that will be used to reconstruct the model in the target zone can be seen in Figure 1 (columns 2-5). The process of optimizing the ML model is continued until the prediction TOC matches the core rock data as reflected by the minimum AAPE average value and the highest correlation coefficient (R^2). The trained and optimized ML model was then validated using another formation in the same well (Gumai organic material-rich shale).

The well log data selected for the TOC value prediction training were analyzed. The analysis is in the form of looking at the comparison between conventional well logs and TOC values measured in the laboratory. The data used to train the ML model, TOC, were highly correlated with RHOB, while moderately associated with DT, ILD, GR, and NPHI.

The five well logs were selected based on their relationship to the TOC values in the core rock, and the five logs will also be used to train the ML model. For example, the estimated ILD of the measured value is influenced by the richness of kerogen in the source rock [25]; DT decreased with increasing TOC [26]; GR is widely used by several researchers in predicting TOC [27]; the more kerogen increases, the RHOB decreases; and NPHI are sensitive to the presence of organic matter even in small amounts and must be corrected for inorganic matter to find out [11]. The explanation above illustrates that the five well logs can be used to develop the TOC model in this research.

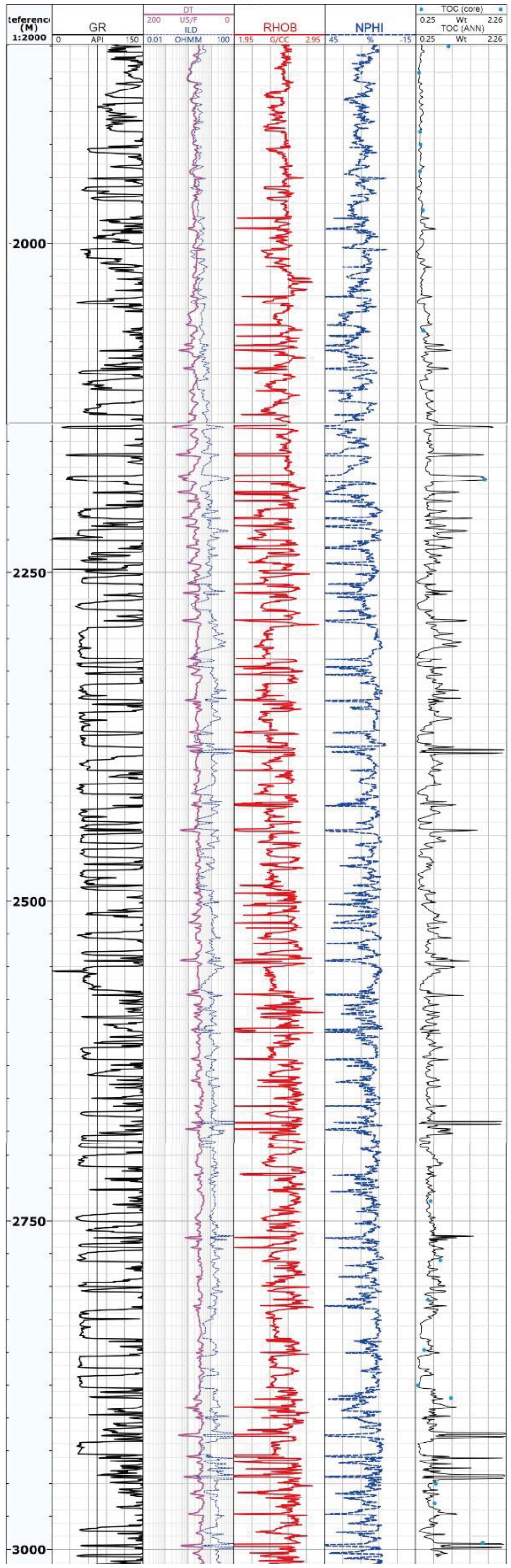


FIGURE 1. Some well log data on the Talang Akar shale formation used to reconstruct the ML model (columns 2-5, start from left) and TOC MLP-ANN prediction compare versus TOC core (column 6).

Principles of MLP-ANN

Multilayer Perceptron (MLP) is a popular supervised learning technique in ANN whose architecture has been used for several forecasting problems in the literature [24]. It is a distributed mathematical model inspired by the behavior of the human brain and nervous system. MLP consists of three layers; input layer, hidden layer, and output layer. The hidden layer may have one or more activation functions [28]. The input for this research is five conventional well logs, while the output is Total Organic Carbon. Figure 2 shows a schematic diagram of the MLP-ANN and the respective neural network architecture.

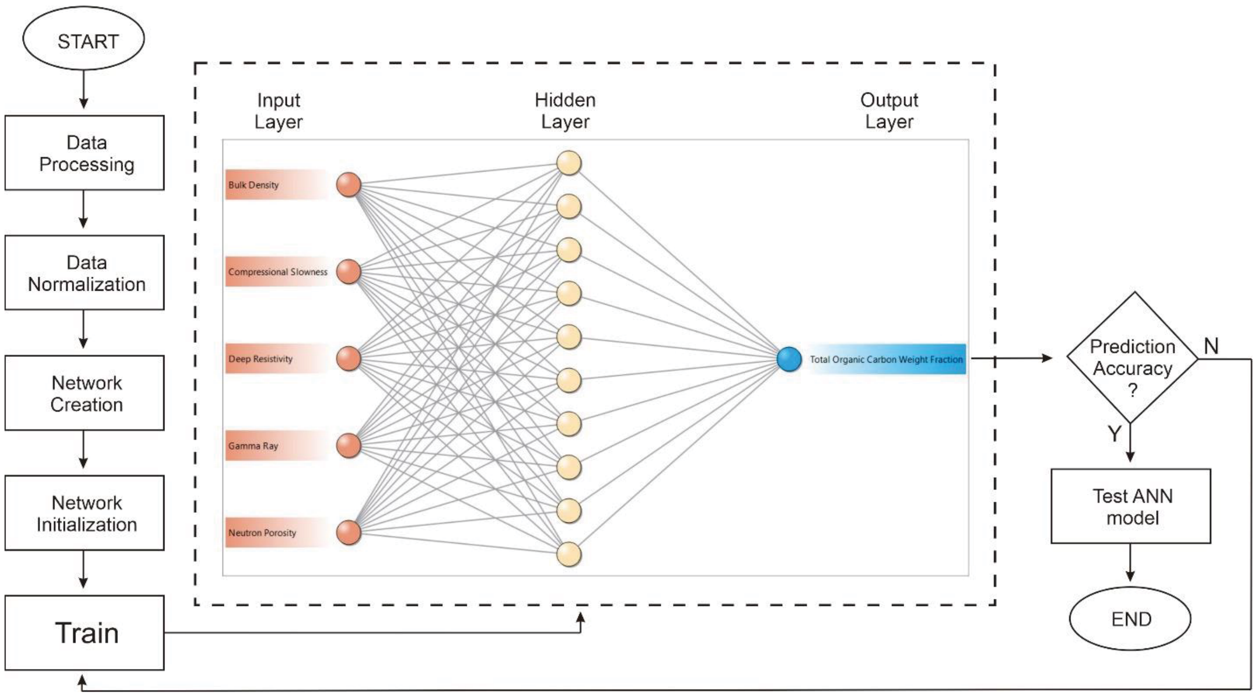


FIGURE 2. The schematic diagram for MLP-ANN.

RESULTS AND DISCUSSION

Figure 1 shows that the MLP-ANN model predicts TOC content with high accuracy with R^2 0.87 and AAPE 10.0% (after validation). Another data set from the same well was used to test the MLP-ANN model using invisible data. TOC values range from 0.3-1.8 Wt%, while RHOB values range from 2.0-2.8 g/cc, and DT values range from 49-124 us/ft. ILD values range from 8-171 ohm.m, then GR values range from 31-173 ^{o}API, and NPHI ranges from 2 to 45%. The test interval zone is the same as the training interval. In Figure 3 it can be seen that the TOC prediction using the MLP-ANN model produces high accuracy (R^2 0.98), especially in the training process. After validation, the MLP-ANN model from several well log data can be concluded that it is feasible to be used to predict TOC. 15% of the data sample is used randomly to assess the generalizability of the reconstructed network so that it is not overfitting. The lowest mean squared error (MSE) shown in Figure 4 is 49, this value indicates the optimal number of iterations. The results obtained at the test stage were analyzed for each evaluation. Network training and performance testing should be appropriate. If only the training stage is seen (eg has a high correlation), then the overfitting effect needs to be considered, and the model is not feasible to choose.

CONCLUSION

In this study, the well log data-based ML model based on the MLP-ANN technique was reconstructed to predict the TOC value, in particular several well logs such as resistivity logs, gamma-rays, transit time, neutron porosity, and density. The model was reconstructed and tested using data on the Talang Akar shale interval zone, which was then validated using data on the Gumai organic material-rich shale interval. The ML model shows a high correlation, both in the Talang Akar and Gumai organic material-rich shale intervals. The model is adequate for estimating TOC using the well logs used in this study (R^2 0.87 and AAPE 10%).

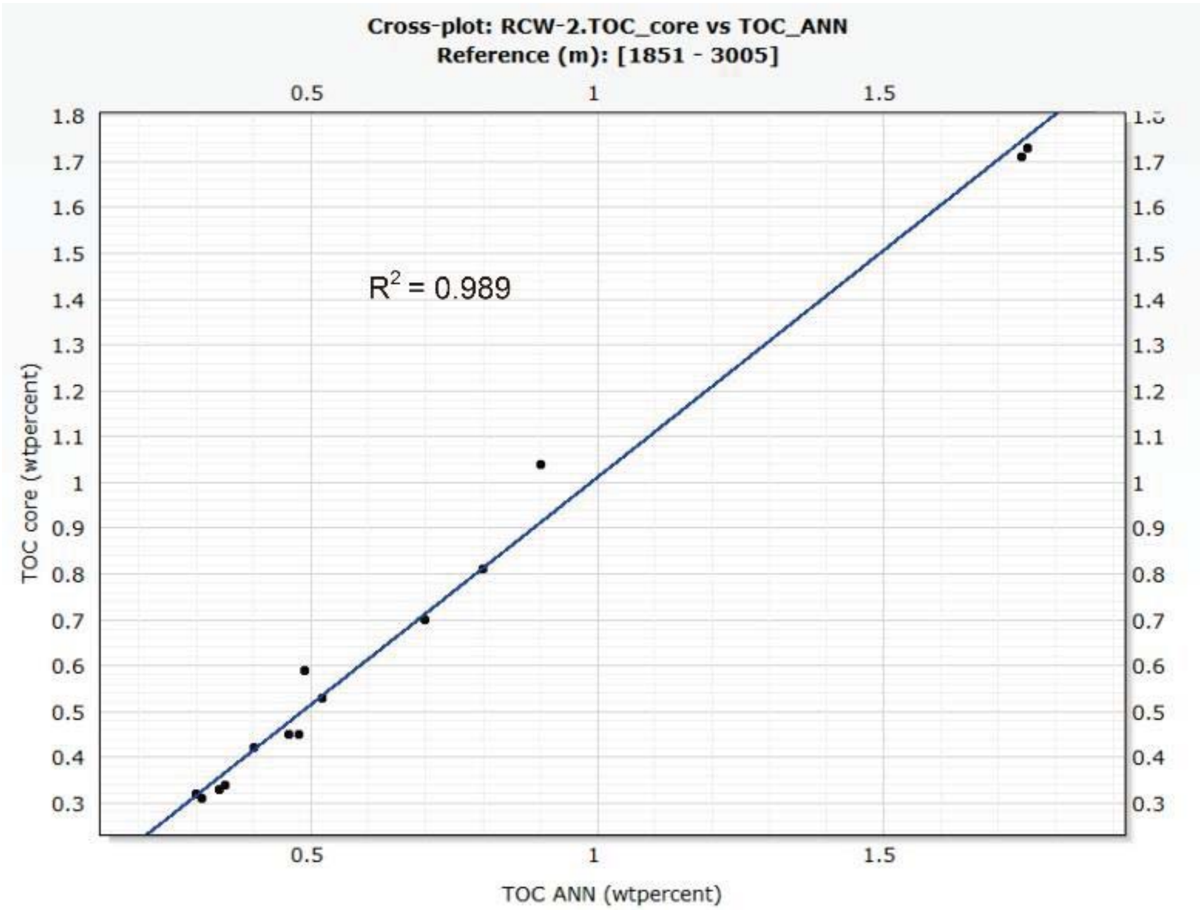


FIGURE 3. The correlation coefficient for the prediction of TOC values at the training stage (18 samples) used the MLP-ANN technique.

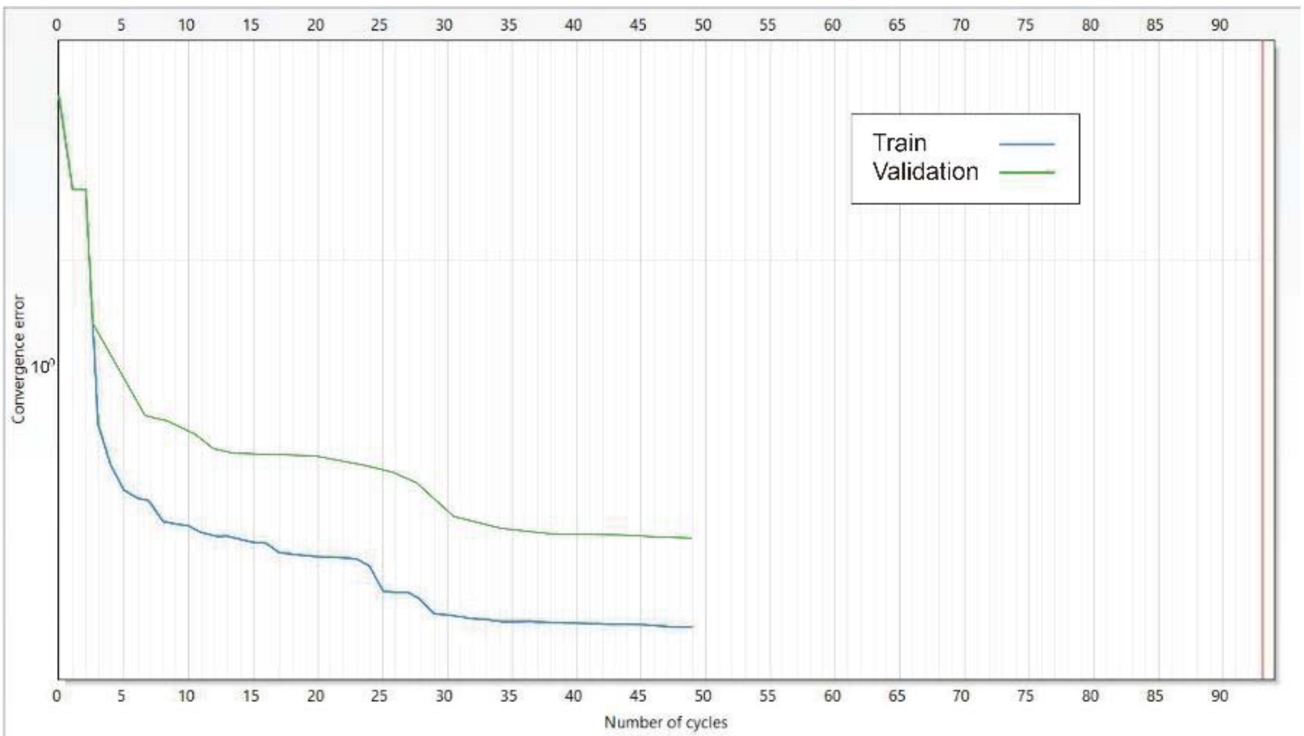


FIGURE 4. Training and validation errors.

ACKNOWLEDGMENTS

We would like to thank the LPPM, the University of Lampung, which has funded this research, and all those who have supported the completion of this article.

REFERENCES

1. Sachsenhofer RF, Leitner B, Linzer HG, Bechtel A, Ćorić S, Gratzer R, et al. Deposition, erosion and hydrocarbon source potential of the Oligocene Eggerding Formation (Molasse Basin, Austria). *Austrian J Earth Sci*., 103(1):76–99, 2010.
2. Bodin S, Fröhlich S, Boutib L, Lahsini S, Redfern J. Early toarcian source-rock potential in the central high atlas basin (central morocco): Regional distribution and depositional model. *J Pet Geol.*, 34(4):345–63, 2011.
3. El Sharawy MS, Gaafar GR. Application of well log analysis for source rock evaluation in the Duwi Formation, Southern Gulf of Suez, Egypt. *J Appl Geophys [Internet]*, 80:129–43, 2012. Available from: http://dx.doi.org/10.1016/j.jappgeo.2011.12.005
4. Amiri Bakhtiar H, Telmadarreie A, Shayesteh M, Heidari Fard MH, Talebi H, Shirband Z. Estimating total organic carbon content and source rock evaluation, applying logr and neural network methods: Ahwaz and marun oilfields, SW of Iran. *Pet Sci Technol*, 29(16):1691–704, 2011.
5. Fertl WH, Rieke HH. Gamma Ray Spectral Evaluation Techniques Identify Fractured Shale Reservoirs and Source-Rock Characteristics. *JPT, J Pet Technol,* 32(11):2053–62, 1980.
6. Fertl WH, Chilingar G V. Total Organic Carbon Content Determined From Well Logs. *SPE Form Eval*, 3(2):407–19, 1988.
7. Schmoker JW, Hester TC. Organic Carbon in Bakken Formation, United States Portion of Williston Basin. *Am Assoc Pet Geol Bull*, 67(12):2165–74, 1983.
8. Mendelzon JD, Toksoz MN. *Source Rock Characterization Using Multivariate Analysis Of Log Data*. Source Rock Charact Using Multivar Anal Log Data. 1985 Jun 17;
9. Carpentier B, Huc AY, Bessereau G. Wireline logging and source rocks estimation of organic carbon by the Carbolog method. *Log Analysis*, 32(3):279–97, 1991.
10. Yun H, Xiang J, Liu Z. Estimation method of organic carbon log and its application in ShengLi Oilfield. *Cejing Jishu/Well Logging Technol*, Jan 1;24(5):372-376+381, 2000.
11. Passey QR, Creaney S, Kulla JB, Moretti FJ, Stroud JD. A Practical Model for Organic Richness from Porosity and Resistivity Logs. *Am Assoc Pet Geol Bull*, 74(12):1777–94, 1990.
12. Kamali MR, Mirshady AA. Total organic carbon content determined from well logs using ΔLogR and Neuro Fuzzy techniques. *J Pet Sci Eng*, 45(3–4):141–8, 2004.
13. Pan Renfang WY and SZ. Geochemical Parameters for Shale Gas Exploration and Basic Methods for Well Logging Analysis [Internet], *China Petroleum Exploration*, Vol. 14, p. 6–9, 2009. Available from: http://www.cped.cn
14. Passey QR, Bohacs KM, Esch WL, Klimentidis R, Sinha S. From Oil-Prone Source Rock to Gas-Producing Shale Reservoir – Geologic and Petrophysical Characterization of Unconventional Shale-Gas Reservoirs. In: *The CPS/SPE International Oil & Gas Conference and Exhibition*, p. 29, 2010.
15. Kadkhodaie-Ilkhchi A, Rahimpour-Bonab H, Rezaee M. A committee machine with intelligent systems for estimation of total organic carbon content from petrophysical data: An example from Kangan and Dalan reservoirs in South Pars Gas Field, Iran. *Comput Geosci*, 35(3):459–74, 2009.
16. Khoshnoodkia M, Mohseni H, Rahmani O, Mohammadi A. TOC determination of Gadvan Formation in South Pars Gas field, using artificial intelligent systems and geochemical data. *J Pet Sci Eng [Internet]*, 78(1):119–30, 2011. Available from: http://dx.doi.org/10.1016/j.petrol.2011.05.010
17. Herron MM, Grau JA, Herron SL, Kleinberg RL, Machlus M, Reeder SL, et al. Total organic carbon and formation evaluation with wireline logs in the green river oil shale. *Proc - SPE Annu Tech Conf Exhib*, 5(October):3709–27, 2011.
18. Radtke RJ, Lorente M, Adolph B, Berheide M, Fricke S, Grau J, et al. A New Capture And Inelastic Spectroscopy Tool Takes Geochemical Logging To The Next Level. *Petrophysics - SPWLA J Form Eval Reserv Descr*, 2012.
19. Tariq Z, Mahmoud M, Abouelresh M, Abdulraheem A. Data-Driven Approaches to Predict Thermal Maturity Indices of Organic Matter Using Artificial Neural Networks. *ACS Omega*, 5(40):26169–81, 2020.
20. Wibowo RC, Mulyatno BS. Karakterisasi Reservoar Menggunakan Metode Inversi Impedansi Akustik dan

Neural Network Pada Lapangan " ICL " Cekungan Sumatera Selatan. In: *PROCEEDINGS PIT HAGI 2012. Palembang: HAGI*, p. 2–4, 2012.

21. Wibowo RC, Sarkowi M, Mulyatno BS, Dewanto O, Zaenudin A, Amijaya DH, et al. Thinned coal distribution modeling based on integrated geological and geophysical data : Case study CBM resources in Central Palembang Sub-Basin. In: *2nd International Conference on Earth Science, Mineral, and Energy. Yogyakarta: AIP Publishing*, p. 1–9, 2020.
22. Rahaman MS., Vasant P. Artificial Intelligence Approach for Predicting TOC From Well Logs in Shale Reservoirs: A Review. In: *IGI Global*, p. 46–77, 2020.
23. Hamada GM, Ahmed E, Chaw Y N. Artificial neural network (ANN) prediction of porosity and water saturation of shaly sandstone reservoirs. *Adv Appl Sci Res [Internet]*, 9(2):26–31, 2018. Available from: www.pelagiaresearchlibrary.com
24. Waszkiewicz S, Krakowska-Madejska P, Puskarczyk E. Estimation of absolute permeability using artificial neural networks (multilayer perceptrons) based on well logs and laboratory data from Silurian and Ordovician deposits in SE Poland. *Acta Geophys [Internet]*, 67(6):1885–94, 2019. Available from: https://doi.org/10.1007/s11600-019-00347-6
25. Heslop KA. Generalized Method for the Estimation of TOC from GR and Rt. In: AAPG Annual Convention and Exhibition. New Orleans; 2010.
26. Liu Y, Chen Z, Hu K, Liu C. Quantifying Total Organic Carbon (TOC) from Well Logs Using Support Vector Regression. In: *GeoConvention 2013*: Integration [Internet], p. 6, 2013. Available from: http://www.geoconvention.com/archives/2013/281_GC2013_Quantifying_Total_Organic_Carbon.pdf
27. Zhao T, Verma S, Devegowda D, Jayaram V. TOC estimation in the Barnett Shale from triple combo logs using support vector machine TOC estimation from triple combo logs using support vector machine. In: *SEG Annual Meeting. New Orleans*, p. 791–5, 2015.
28. Kişi Ö. Multi-layer perceptrons with Levenberg-Marquardt training algorithm for suspended sediment concentration prediction and estimation. *Hydrol Sci J*, 49(6):1025–40, 2004.