Bulletin of Geophysics and Oceanography Vol. XX, n. X, pp. XXX-XXX; Xxxxx 20xx
DOI 10.4430/bgo00523

An artificial neural network model for the prediction of the bulk density log

K. Zabel, M.C. Berguig and R. Ketteb

Geophysics Laboratory, Universite des Sciences et de la Technologie Houari Boumediene, Algiers, Algeria

(Received: 15 May 2024; accepted: 23 December 2025; published online: 6 March 2026)

ABSTRACT Formation evaluation requires the analysis of well-log data that are not available for all wells, especially old wells. In the case of an unstable borehole well, the formation bulk density (RHOB) log cannot reflect the true values of the formation density, and we cannot record it another time. However, the use of artificial intelligence approaches, such as artificial neural networks (ANNs), which are strong tools for real-time prediction without additional costs, can provide acceptable results with high accuracy. This work, by leveraging ANNs for continuous and reliable *in-situ* prediction from other wireline logs, enables the prediction of RHOBs without additional costly logging operations and overcomes the limitations of traditional methods in terms of availability, accuracy, cost, and real-time applicability. The dataset of two wells was used to train and test the model. Unseen data from another well within the same field were utilised for validation. The findings revealed that the predicted RHOB values significantly matched the actual values with a coefficient of determination of 0.98 and root-mean-square error of 0.12. This score confirms the generalisation capability of the model, overcoming, thus, limitations of traditional logging methods that require direct measurement and stable boreholes. This demonstrates the advantage of ANN in filling data gaps and enhancing well-log interpretation under challenging conditions.

Key words: artificial intelligence, artificial neural network, formation evaluation, well logs, formation bulk density.

1. Introduction

A critical property that has an important role in rock interpretation and characterisation is formation bulk density (RHOB) (Carmichael and Klein, 2021). The ratio of the weight of the grains and cementation materials to the total volume of the grains and the pore space defines the RHOB. Because of porosity and mineralogy disparities, rocks have a broad range of RHOBs. Knowing the density distribution of a formation helps identify petrophysical parameters such as porosity (Ellis, 2003), lithology, and fluid content (Alger and Raymer, 1963); geomechanical properties such as overburden stress and pore pressure (Oloruntobi and Butt, 2019); and the Young modulus, bulk modulus, shear modulus, and matrix compressibility (Feng *et al.*, 2019). Accordingly, subsurface geological structure and reservoir characteristics can be interpreted, which can help in drilling optimisation and the development of completion and production strategies (Yusuf *et al.*, 2019). In other cases, the RHOB is frequently affected by wellbore environment effects, such as enlargement or breakouts of the borehole wall, circulation loss, and kicks (de Macedo *et al.*, 2019). Thus, many researchers have proposed linear (Gardner *et al.*, 1974) and nonlinear (Oloruntobi and Butt, 2019) empirical mathematical equations for calculating RHOB from other logs measured in the well (P-wave velocity, porosity, gamma-ray, and other logs). The use of empirical models is limited to one type of lithology, such as clean and compacted shale formations (Yusuf *et al.*, 2019; Gowida *et al.*, 2020), where it is impossible to determine a stable value for the parameters (either of the linear or nonlinear model) that can be applied to wells because the parameter values are affected by different factors that change according to the rock content and the conditions that affect the study area. Thus, geoscientists attempt to generate missed logs on the basis of neighbouring wells where the target log is available via artificial intelligence (AI) algorithms.

The application of AI has been very active in the recent past. Several problems have been addressed in the area of petroleum engineering applications. These include the prediction of the petrophysical and geomechanical parameters. Specifically, artificial neural network (ANN) applications for well-log prediction highlight advances in the use of deep learning models. Salehi *et al.* (2016) predicted missing well logs on the basis of large datasets from several wells via ANNs to improve data completeness and support petrophysical interpretation in well-log analysis. The findings demonstrated effective prediction accuracy, enabling the reconstruction of critical logs. Ghaithi *et al.* (2022) built a deep learning feedforward neural network model to predict shear sonic logs via a large dataset for a carbonate reservoir. The results indicate that almost all the predicted shear logs yielded excellent values. Haritha *et al.* (2025) proved that using a hybrid convolutional neural network (CNN) and bidirectional long short-term memory (Bi-LSTM) (collectively referred to as CNN-Bi-LSTM) model to generate missing well-log data yielded better prediction performance than standalone ANN and Bi-LSTM models did. Al-Fakih *et al.* (2025) achieved promising performance for incomplete log datasets via sequence-based generative adversarial networks for synthetic well-log data generation and imputation. Saleh *et al.* (2025) predicted compressional sonic logs via multiple regression models, including deep neural networks, random forest, CatBoost, and XGBoost. Findings showed the best accuracy, with correlation coefficients (*R*s) near 0.9 and low root-mean-square errors (*RMSE*s) when the ensemble methods were used.

Gowida *et al.* (2019) established an ANN designed to forecast the RHOB in real time while drilling. The inputs for the ANN were derived from mechanical drilling parameters, including the rate of penetration, weight on bit, torque, standpipe pressure, and rotating speed. These parameters are notably influenced by the types of formations encountered and petrophysical well-log data for RHOB as outputs. Ahmed *et al.* (2022) developed various AI models to predict the RHOB of complex lithologies in real time. The ANN and adaptive neuro-fuzzy inference system (ANFIS) techniques are used with the mechanical drilling parameters as inputs. A different dataset from the same tested field was used to validate the developed models. The outcomes obtained demonstrated that both the ANN and the ANFIS-based models estimated the rock density with high fitting accuracy for real-time prediction of the RHOB. Wang *et al.* (2022) applied a CNN-Bi-LSTM for missing well log prediction, demonstrating an improved ability to capture the sequential and spatial features of logs. Wang *et al.* (2023) proposed well-log prediction while drilling by integrating seismic impedance with three neural networks: long short-term memory (LSTM), Bi-LSTM, and a new double chain LSTM (DC-LSTM). Bi-LSTM and DC-LSTM outperform classical LSTM both in accuracy and efficiency, with DC-LSTM reducing training parameters and computation time significantly while maintaining prediction quality. Rahmati *et al.* (2024) focused on predicting RHOBs via machine learning models coupled with heuristic optimisation to improve accuracy. The findings yielded effective and interpretable models for

RHOB prediction from well logs, addressing challenges in reservoir evaluation.

The aforementioned studies highlight the performance of AI in predicting missing logs, specifically RHOB, by means of drilling data and hybrid deep learning methods coupled with heuristic optimisation to tune hyperparameters to improve accuracy. The current work aims to assess the performance of the ANN algorithm in predicting RHOBs from well log such as: compressional wave transit times (DTs) and shear wave transit times (DTSs) and the photoelectric factor (PEF) without incorporating hybrid algorithms, drilling data, or geological information to demonstrate the ability of the neural networks to capture complex nonlinear relationships between standard well logs and RHOB. This approach provides a valuable baseline model for RHOB prediction with practical simplicity and minimal data requirements, supporting its use in streamlined or preliminary well-log analyses.

2. Methodology

ANNs are computational models inspired by the human nervous system and consist of processing units represented as artificial neurons connected by numerous artificial synapses. These connections are established through vectors and matrices of synaptic weights (Raschka and Mirjalili, 2017). During training, the network starts with random weight assignments and strives to align outputs with the desired target values (e.g. the RHOB log). As training advances, output calculations and error assessments occur, and the stopping criterion is met when the error reaches a specified level (Maleki *et al.*, 2022).

The ANN model was structured via Python code. The model was built and evaluated via the Keras sequential application programme interface, which constructs neural networks by linearly stacking layers. The architecture starts with an input layer for three features, followed by four hidden dense (fully connected) layers with 20, 15, 10, and 5 neurons, each using the rectified Linear Unit activation function and normal weight initialisation (Fig. 1). The output layer has a single neuron for regression tasks. The model uses the Adam optimiser and mean-square error loss. To ensure optimal performance, a ModelCheckpoint callback monitors the validation loss and saves the best model weights. Using the Training_test_split function from

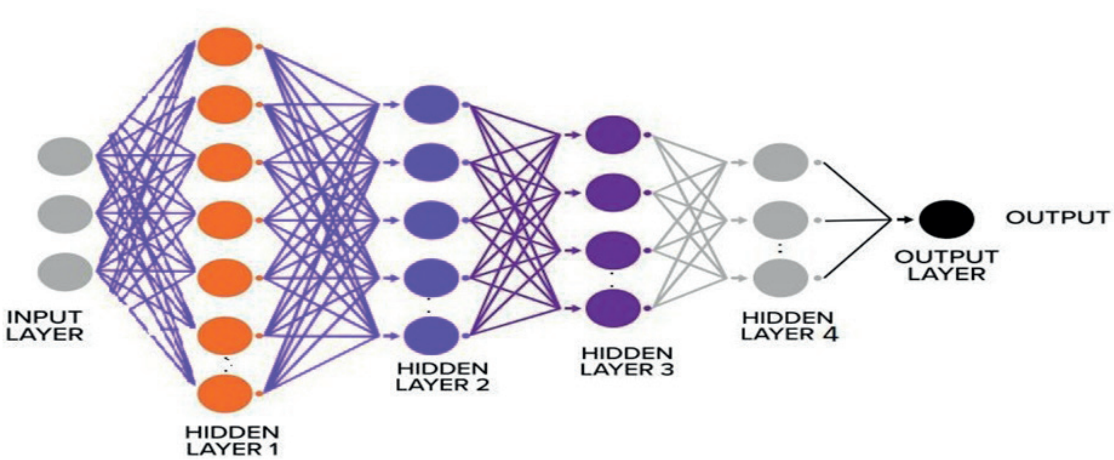

Fig. 1 - The architecture of the developed model.

the scikit-learn library, the data are randomly divided so that 80% is allocated for model training and 20% for testing. This approach enables the resulting model to be trained on one portion, while its generalisation performance is robustly evaluated on previously unseen data. Before developing the ANN model, it is very important to check the effect of each parameter on the required output (Khaled *et al.*, 2022). In other words, the input data are selected on the basis of their relevance in relation to the actual RHOB, which was determined in this study on the basis of the correlation coefficient (R). The findings show that the density log is directly proportional to the PEF and inversely proportional to the sonic log data (Fig. 2). These subsurface physical properties are fundamentally linked to rock composition and structure. Sonic logs (DT and DTS) depend on both lithology and rock density; higher RHOB generally leads to lower travel times for both compressional and shear waves, resulting in strong negative correlations, as seen in the correlation matrix and bar chart. Moreover, the PEF log, which is influenced by mineralogy and RHOB, reflects the electron density of the formation, thus providing complementary information that can be used to distinguish between lithological variations and matrix effects. Moreover, Table 1 shows the statistical parameters, such as count, mean, standard deviation, minimum and maximum values, as well as quartiles for each variable. These statistics provide a clear overview of the central tendency and variability in the combined well data.

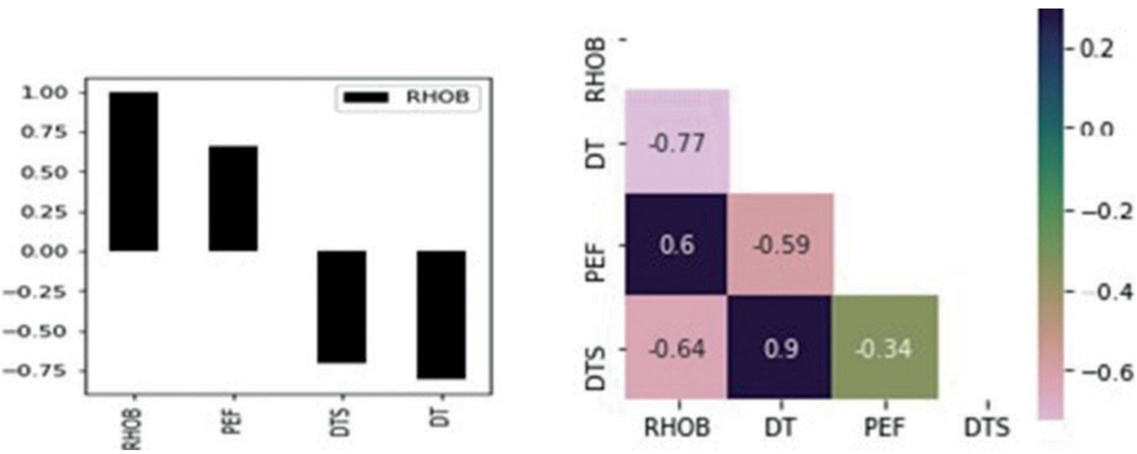

Fig. 2 - The correlation coefficient between bulk density and the well logs parameters (left), heatmap shows the correlation coefficient among input data (right).

To ensure the performance of the algorithm, the data used in this study were processed. On the contrary, the prediction accuracy of the AI-based models would have been significantly affected by the quality of the data used while developing the model. Therefore, the data were preprocessed and filtered via statistical analysis. The preprocessing operation includes normalisation, which is a linear scaling technique. This is performed because the AI algorithms tend to perform better or converge faster when the different features are on a smaller scale. In addition, it makes the training process less sensitive to the scale of the features. This results in better coefficients after training. Therefore, it is common practice to normalise data before training AI models. The Yeo-Johnson power transform is a data normalisation technique that helps stabilise variance and make data distributions more Gaussian-like. In machine learning preprocessing pipelines, it reduces skewness and improves model accuracy when algorithms assume normally distributed features. Next, an attempt was made to remove the outliers (Fig. 3)

Table 1 - Summary of the statistical properties of the dataset used.

	Depth	RHOB	DT	DTS	PEF
Count	13027	13027	13027	13027	13027
Mean	3161.177	2.480	77.361	140.534	6.728
Std	271.453	0.137	12.071	24.491	1.049
Min	2619.800	1.981	56.382	96.901	4.298
25%	2945.450	2.417	69.066	123.942	5.966
50%	3201.900	2.524	74.424	133.882	6.803
75%	3370.050	2.575	83.782	148.487	7.637
Max	3641.800	3.052	116.232	217.970	10.988

(Kim, 2022) via a one-class support vector machine from scikit-learn which performs unsupervised anomaly detection on the training dataset. Furthermore, when coefficient of determination (R^2) and the *RMSE* are measured, the performance of the ANN model is checked:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\sum_{i=1}^{n}(y_i - \bar{y}_i)} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}[f(x_i) - y_i]^2}{n}} \quad (2)$$

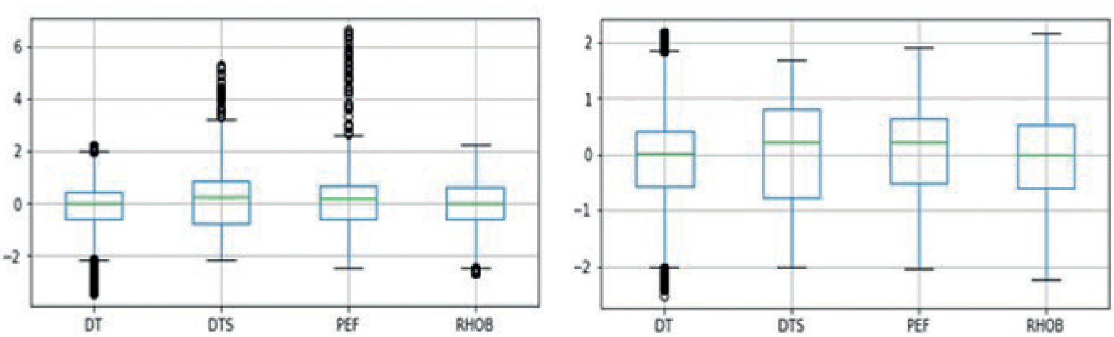

Fig. 3 - Box plot before (left) and after (right) removing outliers.

3. Results and discussion

The results revealed that the predicted RHOB values significantly matched the actual values, with a R^2 value of 0.98 and a *RMSE* of 0.12 (Fig. 4). This score confirms the high accuracy of the prediction of the RHOB values via the developed ANN model. The selection of DTs, DTSs, and PEFs as predictive features, rather than other logs, is justified by their direct physical relationships and significant *Rs* with RHOB, which enhances the accuracy and robustness of data-driven models in petrophysical analysis. The high accuracy of the RHOB values predicted via the developed ANN model was obtained after iterative trial-and-error processes. During model development, an iterative tuning process was used to achieve robust predictive performance. Therefore,

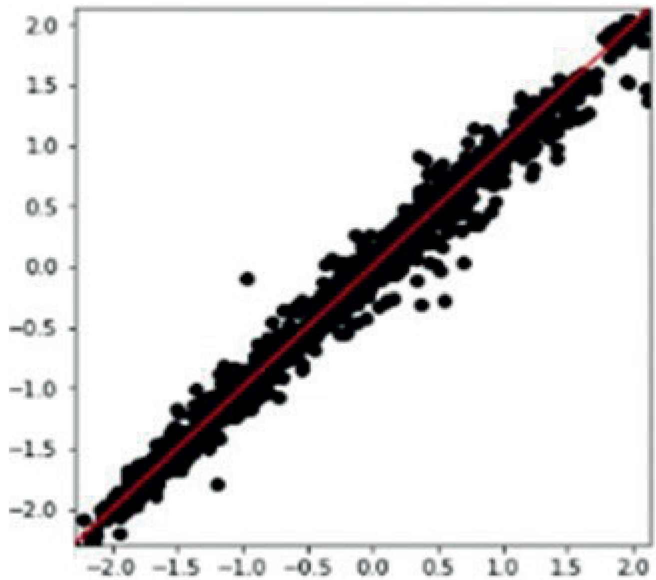

Fig. 4 - True RHOB versus predicted RHOB.

hyperparameters, such as the learning rate, number of hidden layers, number of neurons per layer, and activation functions, were extensively tuned to optimise convergence and minimise error. To avoid overfitting or underfitting and to ensure generalisability, parameter selection was driven by multiple rounds of training, validation, and performance evaluation rather than by the application of arbitrary selection. This result surpasses many conventional empirical models (i.e. the Gardner relation) which typically report R^2 values between 0.85 and 0.95, with higher *RMSE* ranges in similar tasks. Ghareb *et al.* (2018) reported the robustness of multilayer perceptron-based ANN models in capturing nonlinear relationships in reservoir properties and the advantage in generalising from training data. Omidi *et al.* (2022) demonstrated that deep learning methods, including ANNs, outperform traditional tree-based and empirical models, achieving lower *RMSE* values and higher R^2 values for heterogeneous and incomplete data environments. The improved generalisability and nonlinear modelling capability of ANNs make them highly suitable for robust RHOB estimation across varying lithologies and data qualities. Nevertheless, ANNs act as black boxes with limited interpretability, complicating trust and acceptance in high-stake geoscience applications. For this reason, ANN models require large and representative datasets to avoid overfitting, and training, via backpropagation, may slowly converge or become trapped in local minima. The model's limitation appears with novel geological formations where scenarios were underrepresented in the training data. Strengthening the model's robustness and practical applicability in the field requires expanding training data diversity, integrating hybrid models that combine physics-based approaches with data-driven approaches, enhancing interpretability through explainable AI techniques, and adopting optimisation strategies that accelerate convergence and avoid local minima, which represent potential areas of improvement.

4. Conclusions

The purpose of the present work is to generate missing log data via ANN algorithms, which are intended to predict density values without additional costly logging operations by leveraging an ANN for continuous, reliable *in-situ* prediction from other wireline logs and to overcome the limitations of traditional methods in terms of availability, accuracy, cost, and real-time applicability, using three log parameters (DT, DTS, and PEF) collected from two wells and chosen on the basis of the influence of each parameter on the target log.

The results demonstrate that the developed model surpassed the accuracy of density log predictions, achieving an R^2 value of 0.98 and an *RMSE* of 0.12, confirming the reliability of the performance of ANNs for accurately generating density values when well-log data are incomplete or missing and, consequently, reducing the need for expensive and time-consuming additional logging operations and enabling continuous monitoring and real-time decision-making during drilling and production. Moreover, the developed model can lead to better formation evaluation (petrophysical analysis and geomechanical study) and enhanced reservoir modelling. For all these reasons, the expansion of training datasets to include more heterogeneous geological settings and multiple wells, to enhance model generalisability, and the incorporation of hybrid models that combine ANNs with physical or rule-based approaches which can potentially improve prediction robustness and interpretability, should be explored in the future. Finally, extending sensitivity analysis to more complex input suites and testing model performance in real-time operational environments will be valuable for fully establishing the model's utility in the field.

REFERENCES

Ashraf A., Hany G., Salaheldin E. and Abdelwahab A.; 2022: *Bulk density prediction while drilling vertical complex lithology using artificial intelligence.* J. Appl. Geophys., 199, 104574, doi: 10.1016/j.jappgeo.2022.104574.

Al-Fakih E., Al-Mousa K. and Al-Sabah K.; 2025: *Well log data generation and imputation using sequence-based generative adversarial networks.* Sci. Rep., 15, 95709, doi: 10.1038/s41598-025-95709-0.

Alger R.P. and Raymer Jr. L.L.; 1963: *Formation density log applications in liquid-filled holes.* J. Pet. Technol., 15, 321-332, doi: 10.2118/435-PA.

Carmichael R.S. and Klein C.; 2021: *Rock.* Encyclopedia Britannica, https://www.britannica.com/science/rock-geology, accessed 15 February 2021.

Ellis D.; 2003: *Formation porosity estimation from density logs*. Petrophys. 44, 306-316.

Feng C., Wang Z., Deng X., Fu J., Shi Y., Zhang H. and Mao Z.; 2019: *A new empirical method based on piecewise linear model to predict static Poisson's ratio via well logs.* J. Pet. Sci. Eng., 175, doi: 10.1016/j.petrol.2018.11.062.

Gardner G.H.F., Gardner L.W. and Gregory A.R.; 1974: *Formation velocity and density; the diagnostic basic for stratigraphic traps.* Geophys., 39, 770-780, doi: 10.1190/1.1440465.

Ghaithi A.Al and Aranco S.; 2022: *A robust approach for shear log predictions using deep learning an big dataset from a carbonate reservoir for integrated reservoir characterization project.* In: Second International Meeting for Applied Geoscience & Energy, pp. 1840-1844, doi: 10.1190/image2022-3746877.1.

Gowida A., Elkatatny S. and Abdulraheem A.; 2019: *Application of artificial neural network to predict formation bulk density while drilling.* Petrophys., 60, 660-674, doi: 10.30632/PJV60N5-2019a9.

Haritha D., Satyayani N. and Ramesh A.; 2025: *Generation of missing well log data with deep learning: CNN-Bi-LSTM approach.* J. App. Geophys., 233, 105628, doi: 10.1016/j.jappgeo.2025.105628.

Khaled S., Soliman A.A., Mohamed A., Gomaa S. and Attia A.M.; 2022: *New models for predicting pore pressure and fracture pressure while drilling in mixed lithologies using artificial neural networks.* ACS Omega, 7, 31691-31699, doi: 10.1021/acsomega.2c01602.

Kim J.; 2022: *Synthetic shear sonic log generation utilizing hybrid machine learning techniques.* Artif. Intell. Geosci., 3, 53-70, doi: 10.1016/j.aiig.2022.09.001.

Maleki S., Ramazi H.R., Javad M. and Shahrabi A.; 2022: *A new 3D model for shear wave velocity by utilizing conventional petrophysical logs and geostatistical method.* J. Min. Environ., 13, 431-447, doi: 10.22044/jme.2022.11462.2134.

Oloruntobi O. and Butt S.; 2019: *The new formation bulk density predictions for siliciclastic rocks.* J. Pet. Sci. Eng., 180, 526-537.

Omidi S., Arasteh M. and Hashemi S.; 2022: *Using an artificial neural network model to predict the number of COVID-19 cases.* Health Data Q., 17, 456-467, doi: 10.1016/j.hdq.2022.03.001.

Rahmati M., Zargar G. and Tanha A.; 2024: *Predicting density log from well log using machine learning techniques and heuristic optimization algorithm: a comparative study.* Pet. Res., 9, 176-192, doi: 10.1016/j.ptlrs.2024.01.008.

Raschka S. and Mirjalili V.; 2017: *Machine learning and deep learning with python, scikit-learn, and tensorflow.* Packt Publishing Ltd., Birmingham, UK, 339 pp.

Saleh A., Elsayed S., Abdelaziz T. and Elbeltagi E.; 2025: *Machine learning model optimization for compressional sonic log prediction using multiple regression and ensemble methods*. Sci. Rep., 15, 97938, doi: 10.1038/s41598-025-97938-9.

Salehi M., Rahmati M., Karminezhad M. and Omidvar P.; 2016: *Estimation of the non records logs from existing logs using artificial neural networks.* Egypt. J. Pet., 26, 957-968, doi: 10.1016/j.ejpe.2016.11.002.

Wang H., Xu Y., Tang S., Wu L., Cao W. and Huang X.; 2023: *Well log prediction while drilling using seismic impedance integrated with LSTM, biLSTM, and double chain LSTM models.* Front. Earth Sci., 11, 1153619, doi: 10.3389/feart.2023.1153619.

Corresponding author: Khalda Zabel
Geophysics Laboratory, Universite des Sciences et de la Technologie Houari Boumediene
Bab-Ezzouar, BP 32 El Alia, 16111 Algiers, Algeria
Phone: +213 0541354840; e-mail: khalda.zabel@usthb.dz