

## Density Well Log Prediction in X Field Niger Delta using Ensemble Learning Models and Artificial Neural Network

Patient K. Mulekya <sup>1,2</sup>, Olugbenga A. Boboye <sup>1,3</sup>, Moruffdeen A. Adabanija <sup>1,4</sup>,  
Kasongo Numbi <sup>2</sup>, Tomisin B. Babalola <sup>1</sup>

<sup>1</sup>Petroleum Geoscience Program, Pan African University Life and Earth Sciences Institute, University of Ibadan, Ibadan, Nigeria

<sup>2</sup>Department of Exploration and Production, University of Kinshasa, Kinshasa, DR Congo

<sup>3</sup>Department of Geology, University of Ibadan, Ibadan, Nigeria

<sup>4</sup>Department of Earth Sciences, Ladoko Akintola University of Technology, Ogbomoso, Nigeria

\* Correspondence: [patient.mulekya@unikin.ac.id](mailto:patient.mulekya@unikin.ac.id)

Received: 14 April 2024 / Accepted: 02 August 2024 / Published: 20 August 2024

**Abstract:** Performing reservoir characterization in exploration with limited data can be very challenging. Various approaches are used to estimate values away from the well location. This study addresses the missing density log (RHOB) in one of the five available well log datasets, which is crucial for porosity analysis. An artificial neural network (ANN) approach was employed to synthesize a density log using the available Gamma Ray (GR) log, Sonic (DT), water saturation (SW), and Depth data from three wells in the field. The performance of the prediction was evaluated using the fourth well. Five models were constructed with different optimizers: Nesterov-accelerated Adaptive Moment Estimation (NADAM), Adaptive Moment Estimation (ADAM), Stochastic Gradient Descent (SGD), and Root Mean Square Propagation (RMSP), as well as an Ensemble model combining the four optimizers. Tests on actual data revealed very low mean absolute errors of 0.0262, 0.0278, 0.0270, 0.0309, and 0.0248, and high coefficients of determination ( $R^2$ ) of 0.8832, 0.8746, 0.8986, 0.8858, and 0.9051 for NADAM, ADAM, SGD, RMSP, and the Ensemble model, respectively. These results highlight the high performance of the Ensemble Learning model, suggesting its effectiveness for predicting the missing RHOB log.

**Keywords:** RHOB, Ensemble learning, ANN, Niger Delta

### INTRODUCTION

Artificial neural networks (ANNs) have revolutionized the field of machine learning by mimicking the complex processes of the human brain. These computational models are designed to replicate the learning mechanisms of biological neural systems, leveraging experience and observation to identify patterns and make predictions (Lim, 2005; Miri et al., 2007). ANNs consist of interconnected nodes or neurons that process information through weighted connections, and they are organized into layers: an input layer, one or more hidden layers, and an output layer. Each neuron in the network receives inputs, applies transformations via an activation function, and produces an output that is passed to other neurons. This process is iterative, with the network continuously adjusting the weights of connections during training to minimize the error between predicted and actual outputs (Aggrey & Davies, 2007).

The fundamental strength of ANNs lies in their ability to model complex and nonlinear relationships within data. This capability is achieved through nonlinear activation functions, such as sigmoid, tanh, or ReLU, which introduce nonlinearity into the network's computations (Saedi et al., 2007). By stacking multiple layers of neurons and applying these activation functions, ANNs can capture intricate patterns and interactions in data, making them highly effective for tasks that involve complex data structures (Ozbayoglu & Ozbayoglu, 2007).

ANNs have demonstrated remarkable versatility and effectiveness across various domains, including image recognition, natural language processing, and financial forecasting (Tamhane et al., 2000; Mohaghegh, 2000; Al-Bulushi et al., 2009; Chen et al., 2017). One of the significant areas where ANNs have made a profound impact is in reservoir characterization within petroleum geoscience. Reservoir characterization involves integrating diverse data sources, such as geological, petrophysical, and geophysical data, to understand and predict subsurface conditions (Saikia et al.,

2020). Accurate characterization is essential for optimizing reservoir management and improving hydrocarbon recovery.

Porosity and permeability are two critical properties in reservoir characterization. Porosity measures the volume of void spaces within a rock, while permeability quantifies the ability of fluids to flow through the rock matrix. Accurate estimates of these properties are crucial for designing effective extraction strategies and maximizing resource recovery. However, acquiring precise data for porosity and permeability can be challenging due to various factors, including financial constraints, operational difficulties, and borehole instability.

Density logs are valuable tools for measuring bulk density and porosity in wells. These logs provide essential information about the rock matrix and fluid content within the reservoir. Despite their importance, density logs may sometimes be missing or incomplete due to practical challenges during data acquisition. This can result in gaps within the dataset and negatively affect the accuracy of reservoir models.

To address the issue of missing or incomplete density logs, researchers have explored various data imputation and prediction techniques. ANNs offer a promising approach for synthesizing density logs by leveraging available information from other well logs and measurements. By identifying patterns and correlations within the data, ANNs can generate pseudo-density logs that approximate the missing information. Previous studies have highlighted the effectiveness of ANNs in predicting missing data and enhancing the reliability of reservoir characterization (Adhari & Kardawi, 2022; Abuh et al., 2023; Tugwell & Livinus, 2023; Origbo & Mbachu, 2024).

In addition to ANNs, deep neural networks (DNNs) have been employed to generate synthesized density logs. DNNs, characterized by their multiple hidden layers, offer enhanced capabilities for modeling complex relationships and capturing detailed patterns in the data (Long et al., 2016; Kim et al., 2020). By integrating data from Gamma Ray (GR) logs, Sonic (DT) logs, water saturation (SW) data, and depth information, DNNs can provide valuable estimates of density logs that complement existing datasets.

This study aims to advance the application of ANNs and ensemble learning techniques in predicting density logs for the X field in the Niger Delta, Nigeria. The X field is characterized by its complex geological and petrophysical properties, which present unique challenges for reservoir characterization. To tackle the issue of missing RHOB log data, this research employs four distinct ANN models with different optimizers and integrates them into an ensemble model. The ensemble approach seeks to improve prediction accuracy and robustness by combining the strengths of multiple individual models.

Ensemble learning techniques, such as bagging, boosting, and stacking, aggregate the predictions of various models to enhance overall performance. By leveraging the strengths of different models and mitigating their individual weaknesses, ensemble methods offer a powerful way to improve prediction accuracy and reliability (Chen et al., 2017). In this study, the ensemble approach involves training multiple ANN models with various optimization strategies and combining their predictions to achieve a more precise estimate of density logs.

Through the integration of advanced machine learning techniques, this study aims to enhance the accuracy of porosity and gas level models for the X field reservoir. The application of ensemble learning and ANNs represents a significant advancement in addressing the challenges associated with missing density logs and incomplete datasets in reservoir characterization. The findings from this research have the potential to provide valuable insights into reservoir management and contribute to more effective extraction strategies.

By addressing the critical issue of missing RHOB log data and employing state-of-the-art techniques, this study strives to improve the precision of reservoir models and support informed decision-making in the field of petroleum geoscience. The enhanced predictions and insights gained from this research will contribute to optimizing reservoir management and maximizing resource recovery, ultimately advancing the field of reservoir characterization and management.

### Geological Setting

The X field is situated in the offshore Niger Delta Basin, a prominent geological feature located in the Gulf of Guinea. This basin represents a classic example of a Tertiary delta, characterized by its complex sedimentary sequences and rich hydrocarbon resources. The Niger Delta is known for its substantial accumulation of marine and deltaic clastics, which form a regressive sequence of sedimentary layers. These layers are divided into three primary lithostratigraphic units: the Akata Formation, the Agbada Formation, and the Benin Formation (Adegoke et al., 2017).

At the base of this sedimentary sequence lies the Akata Formation, which consists predominantly of marine shales. This formation represents the initial phase of sediment accumulation in the delta and serves as an important source rock for hydrocarbon generation. Overlying the Akata Formation is the Agbada Formation, a paralic unit characterized by a mixture of deltaic and marine clastics. The Agbada Formation is of significant interest in hydrocarbon exploration as it contains many of the delta's major reservoirs. The formation extends from the Eocene to the Pleistocene epochs and consists of over 3,700 meters of paralic siliciclastics deposited in a deltaic environment during a dominantly regressive phase that started in the mid-Permian and continued until the early Triassic (Tuttle et al., 1999). This extensive sequence of sediments includes a variety of sandstones and shales that are critical to the understanding of reservoir quality and distribution.

The uppermost unit in this stratigraphic sequence is the Benin Formation, which comprises non-marine alluvial continental sands. This formation represents the final phase of sedimentation in the delta, characterized by a transition from deltaic to continental depositional environments. The Benin Formation often serves as a cap rock, providing crucial sealing conditions for the underlying hydrocarbon reservoirs.

The Niger Delta Basin is further divided into several depobelts, which are characterized by progressively younger sedimentary deposits moving outward from the basin center. This structural division reflects the complex tectonic and depositional history of the basin, which includes multiple phases of rifting, subsidence, and sedimentation. The Agbada Formation, in particular, is noted for its complex structural framework, which includes a series of rollover anticlinal traps that are favorable for hydrocarbon accumulation (Ogbamikhumi & Omorogieva, 2021).

The Tertiary Niger Delta (Agbada-Akata) Petroleum System is recognized as the primary petroleum system in the region. This designation is based on the significant oil production derived from the Agbada Formation, with additional contributions from the deeper offshore areas of the Akata Formation. The Agbada Formation is renowned for its prolific hydrocarbon reservoirs, which are hosted in interbedded sands that are encased within these anticlinal traps.

However, the geological complexity of the Niger Delta introduces several challenges for hydrocarbon exploration and production. The presence of heterogeneous lithologies, as reported by Maju-Oyovwikowhe & Osayande (2023), can lead to variations in reservoir properties and complicate the interpretation of well log data. Additionally, the region is characterized by intricate faulting and structural deformation (Fagbemi et al., 2024), which can impact the consistency and reliability of well log measurements.

In the context of these geological challenges, the study area in the X field presents unique difficulties in obtaining consistent and reliable well log data. The heterogeneous nature of the lithologies and the complex faulting patterns can lead to gaps or inconsistencies in the available well log datasets, particularly affecting the measurement of petrophysical properties such as porosity and permeability.

To address these challenges, this study employed an artificial neural network (ANN) approach to synthesize missing density logs (RHOB) from the available well log data. The ANN model leverages the nonlinear relationships among various well log measurements—such as Gamma Ray, Sonic, water saturation, and depth—to predict the missing density log with greater accuracy. By incorporating the specific geological characteristics of the X field, including the lithological heterogeneity and structural complexities inherent to the Niger Delta reservoirs, the ANN models aim to enhance reservoir characterization. This approach provides a more comprehensive and reliable dataset for the development of accurate porosity and gas level models, which are essential for effective reservoir management and resource optimization.

The integration of ANN techniques into the reservoir characterization process addresses the gaps and inconsistencies in the well log data, facilitating improved understanding and modeling of the reservoir properties. This methodology offers significant advantages in dealing with the geological complexities of the Niger Delta, ultimately contributing to more precise and reliable assessments of hydrocarbon reservoirs in the X field.

## METHOD

### Available Data

The study utilized well logs from five wells in the X field: TMG\_2, TMG\_3, TMG\_4, TMG\_8, and TMG\_9 (Table 1). These well logs are essential for comprehensive reservoir characterization and include various types of data:

- Caliper (CALI): Measures the diameter of the borehole, which helps in identifying variations in borehole size and potential formation damage.
- Gamma Ray (GR): Indicates the concentration of natural radioactivity in the formation, used to differentiate between shales and non-shales.
- Spontaneous Potential (SP): Measures the electric potential between the borehole and formation, aiding in identifying reservoir zones.
- Resistivity (RES): Assesses the formation's ability to resist electrical flow, providing insights into fluid saturation and lithology.
- Density (RHOB): Provides bulk density measurements of the formation, crucial for calculating porosity.
- Sonic (DT): Measures the time it takes for a sonic wave to travel through the formation, used to determine formation porosity and mechanical properties.
- Water Saturation (SW): Indicates the fraction of pore volume filled with water, critical for understanding hydrocarbon saturation.

Table 1. Available Well Log Data

Well	CALI	GR	SP	RES	RHOB	DT	SW
TMG_2	✓	✓	✓	✓	✓	✓	✓
TMG_3	✓	✓	✗	✓	✓	✓	✓
TMG_4	✓	✓	✗	✓	✓	✓	✓
TMG_8	✓	✓	✓	✓	✓	✓	✓
TMG_9	✓	✓	✗	✓	✗	✓	✓

### Data Preparation

Data preparation is a crucial step in developing an artificial neural network (ANN) model, as it involves transforming raw data into a format suitable for analysis and modeling (Figure 1).

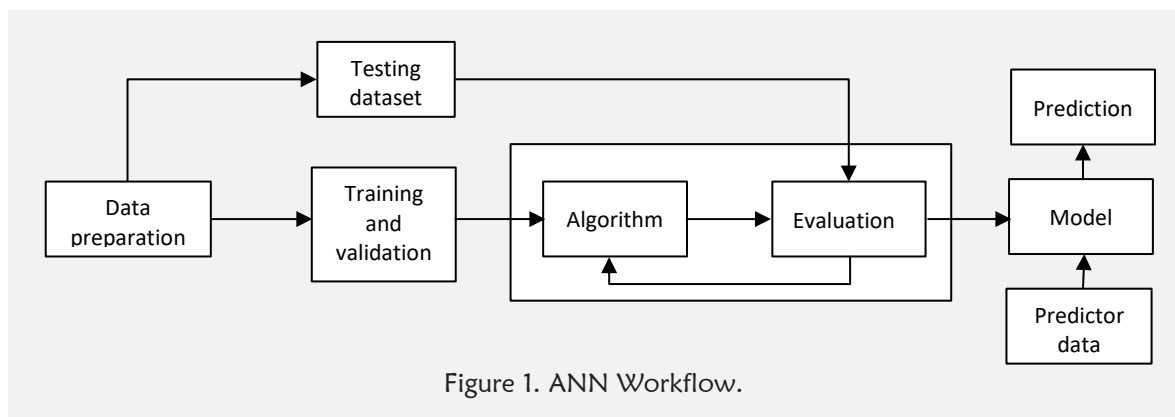


Figure 1. ANN Workflow.

### Removal of Null Data and Outliers

To ensure the integrity of the ANN training process, null values in the dataset were removed. This was necessary to avoid errors that could arise from incomplete data. Outliers were identified using the Z-score method and were clipped to the maximum or minimum values to prevent distortion of the model's performance. Outliers can significantly impact the training process, leading to skewed results, hence their management is crucial for accurate predictions.

### Feature Selection

Selecting features that provide the most information is important for building a good prediction model. This decreases the amount of time needed to train neural networks, reduces the risk of overfitting, and significantly enhances the model's performance. Using the best features rather than all of them speeds up the process and improves the estimation of each model. This selection includes all the variables that may be used to achieve the most accurate performance estimate (Rajabi et al., 2021). Feature selection enhances interpretability by removing redundant and irrelevant variables (Saporetti et al., 2021). A feature selection approach using correlation analysis is commonly applied

to choose pertinent features crucial to the desired output outcomes. This enables the removal of some ineffective variables, improving the efficacy of the analysis (Liu & Liu, 2021).

Based on Pearson's correlation coefficient, four well log variables were retained for training and prediction: Depth, GR, DT, and SW. Table 2 shows the correlation analysis for all the well log variables. RHOB was highly correlated with Depth (0.5), DT (-0.5), GR (0.61), and SW (0.75), which would enhance the model's performance, while CALI (-0.22) and RES (-0.19) would provide the least or biased information and were therefore discarded.

Table 2. Pearson's Correlation Analysis of Well Log Variables

	DEPTH	CALI	DT	RES	GR	SW	RHOB
DEPTH	1	-0.38	-0.7	0.00076	0.082	0.45	0.5
CALI	-0.38	1	0.22	-0.031	-0.15	-0.22	-0.22
DT	-0.7	0.22	1	0.004	-0.08	-0.41	-0.5
RES	0.00076	-0.031	0.004	1	-0.14	-0.078	-0.19
GR	0.082	-0.15	-0.08	-0.14	1	0.45	0.61
SW	0.45	-0.22	-0.41	-0.078	0.45	1	0.75
RHOB	0.5	-0.22	-0.5	-0.19	0.61	0.75	1

### Data Splitting

The prepared data was divided into three subsets: the training dataset, the validation dataset, and the test dataset. This division allows for effective model development and evaluation:

- Training Dataset: Used to train the ANN model, teaching it to recognize patterns and relationships in the data.
- Validation Dataset: Employed to evaluate the model during training, adjusting hyperparameters to optimize performance and prevent overfitting.
- Test Dataset: Used to assess the model's performance on unseen data, providing a measure of how well the model generalizes to new inputs.

For this study, data from TMG\_2, TMG\_3, and TMG\_8 wells were randomly selected for training (80%) and validation (20%) datasets. The test dataset comprised all available data from the TMG\_4 well, while the final prediction of the density log was performed for TMG\_9, which had missing RHOB data.

### Data Normalization

Normalization is crucial for ensuring that all features contribute equally to the model training process. It scales features to a common range, preventing any single feature from disproportionately influencing the model due to its scale. The Z-score normalization method was applied to all input data, calculated as follows:

$$Z_{score} = \frac{x - \bar{x}}{std} \quad (1)$$

where:  $x$  is the original datum;  $\bar{x}$  is the average of the dataset;  $std$  is the standard deviation. Normalization ensures that the different input features contribute equally to the training process, preventing any single feature from dominating due to its scale.

### Model Training

#### ANN structure

The ANN architecture for this study consisted of:

- Input Layer: 4 neurons, corresponding to the four selected features (Depth, GR, DT, SW).
- Hidden Layer: 32 neurons, providing a balance between model complexity and training efficiency.
- Output Layer: 1 neuron, outputting the predicted density log value.

#### Hyperparameter Optimization

Hyperparameter optimization is crucial in developing efficient and accurate models, as it directly influences the model's ability to learn and generalize from the data (Ahmadi & Chen, 2020).

In practical applications, tuning hyperparameters from their default settings is essential to build models that can effectively solve specific problems. This process helps achieve optimal performance by selecting the best configuration of hyperparameters, such as learning rate, epochs, and batch size, tailored to the dataset and the task at hand.

In this study, hyperparameters were optimized using a random search method, which systematically explores a wide range of hyperparameter values to identify the most effective combination. The epoch count and mini-batch size were set to 25 and 8, respectively, based on the results of the optimization process. Additionally, the learning rate for the optimizers was fine-tuned to 0.01, striking a balance between rapid convergence and the prevention of overshooting during training.

### Model Compilation

The performance or accuracy of a statistical model is the primary metric by which it is evaluated. To achieve high accuracy, it's essential to measure the discrepancy between the predicted and actual values at each iteration of model training. This is where loss functions play a critical role, as they quantify the difference between the estimated and actual values, guiding the optimization process to minimize errors. In this study, we employed four different optimizers and two loss functions, which were iteratively adjusted during the model compilation phase to refine the model's predictions.

#### 1. Optimizers

Optimizers are algorithms that adjust the parameters of the neural network, such as the learning rate and weights, to minimize the loss and enhance the model's predictive accuracy. The model's structure is fundamentally shaped by the choice of optimizer. In this study, we selected four popular optimizers:

1. Nesterov-accelerated Adaptive Moment Estimation (NADAM): A variant of the ADAM optimizer that includes Nesterov momentum, providing improved convergence rates in certain scenarios.
2. Adaptive Moment Estimation (ADAM): Widely used due to its adaptive learning rates and ability to handle sparse gradients, making it robust across various types of data.
3. Stochastic Gradient Descent (SGD): A simple yet powerful optimizer that updates the model parameters based on a small batch of training data, contributing to more stable and generalized models.
4. Root Mean Square Propagation (RMSP): An adaptive learning rate method that normalizes the gradients, preventing the learning rate from diminishing too quickly.

These optimizers were evaluated individually, and their performance scores were combined into an Ensemble model using a weighted average. This approach capitalizes on the strengths of each optimizer, leading to a more robust and accurate prediction model.

#### 2. Loss functions

Loss functions are essential for determining how well the model's predictions align with the actual data. They measure the error between predicted and actual values and guide the optimizer in adjusting the model parameters to minimize this error. In this study, we utilized two loss functions:

1. Mean Absolute Error (MAE): This loss function calculates the average of the absolute differences between the actual values ( $x_i$ ) and the predicted values ( $y_i$ ). The MAE is expressed as:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (2)$$

where  $n$  is the total number of data points. MAE is particularly useful when all errors are equally important.

2. Mean Squared Error (MSE): This loss function takes the average of the squared differences between the actual and predicted values. It is expressed as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2 \quad (3)$$

MSE is sensitive to larger errors, making it useful when larger deviations from the actual values need to be penalized more heavily.

By using both MAE and MSE, we ensured that the model is optimized to reduce both the average error and the impact of outliers, resulting in a more accurate and reliable prediction model.

### Performance Metrics and Test

To evaluate the performance of the model, various metrics were employed to compare the actual density log data in the test dataset with the predicted values generated by each model. The selected metrics provide a comprehensive understanding of the model's accuracy and reliability:

1. Root Mean Square Error (RMSE):

RMSE is a widely used metric that quantifies the differences between predicted and actual values by computing the square root of the average of the squared differences. It effectively captures the magnitude of the prediction errors, with a lower RMSE indicating better model performance. RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (4)$$

where:  $y_i$  represents the predictive values;  $x_i$  represents the actual values; and  $n$  is the total number of data points.

2. R-Squared ( $R^2$ ) Coefficient or Coefficient of Determination:

The  $R^2$  coefficient measures the proportion of variance in the actual data that is predictable from the model's predictions. It evaluates the model's strength by comparing the actual density log data in the test dataset to the predicted values. An  $R^2$  value closer to 1 indicates a strong model that explains most of the variance in the data, while a value closer to 0 indicates a weaker model.  $R^2$  is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (x_i - \hat{x})^2} \quad (5)$$

where:  $y_i$  represents the predicted values;  $x_i$  represents the actual values;  $\hat{x}$  is the mean of the actual values; and  $n$  is the total number of data points.

### Prediction

Following the testing, validation, and comprehensive evaluation of performance metrics on the test dataset, the Ensemble model emerged as the most accurate and reliable compared to the other models. Due to its superior performance, the Ensemble model was selected for the crucial task of predicting the missing density log (RHOB) in the TMG-09 well.

The TMG-09 well initially lacked a complete density log, which is essential for accurate porosity analysis and subsequent reservoir characterization. By leveraging the strength of the Ensemble model, which effectively integrates the predictive power of multiple optimizers, the missing RHOB values were successfully predicted. This approach not only restored the incomplete dataset but also enhanced the overall accuracy of reservoir characterization in the X field.

The predicted RHOB log for TMG-09 was then integrated with the existing well log data, enabling a more comprehensive analysis of the reservoir's properties. This prediction was instrumental in filling critical data gaps, thus supporting more informed decision-making in the field development strategy.

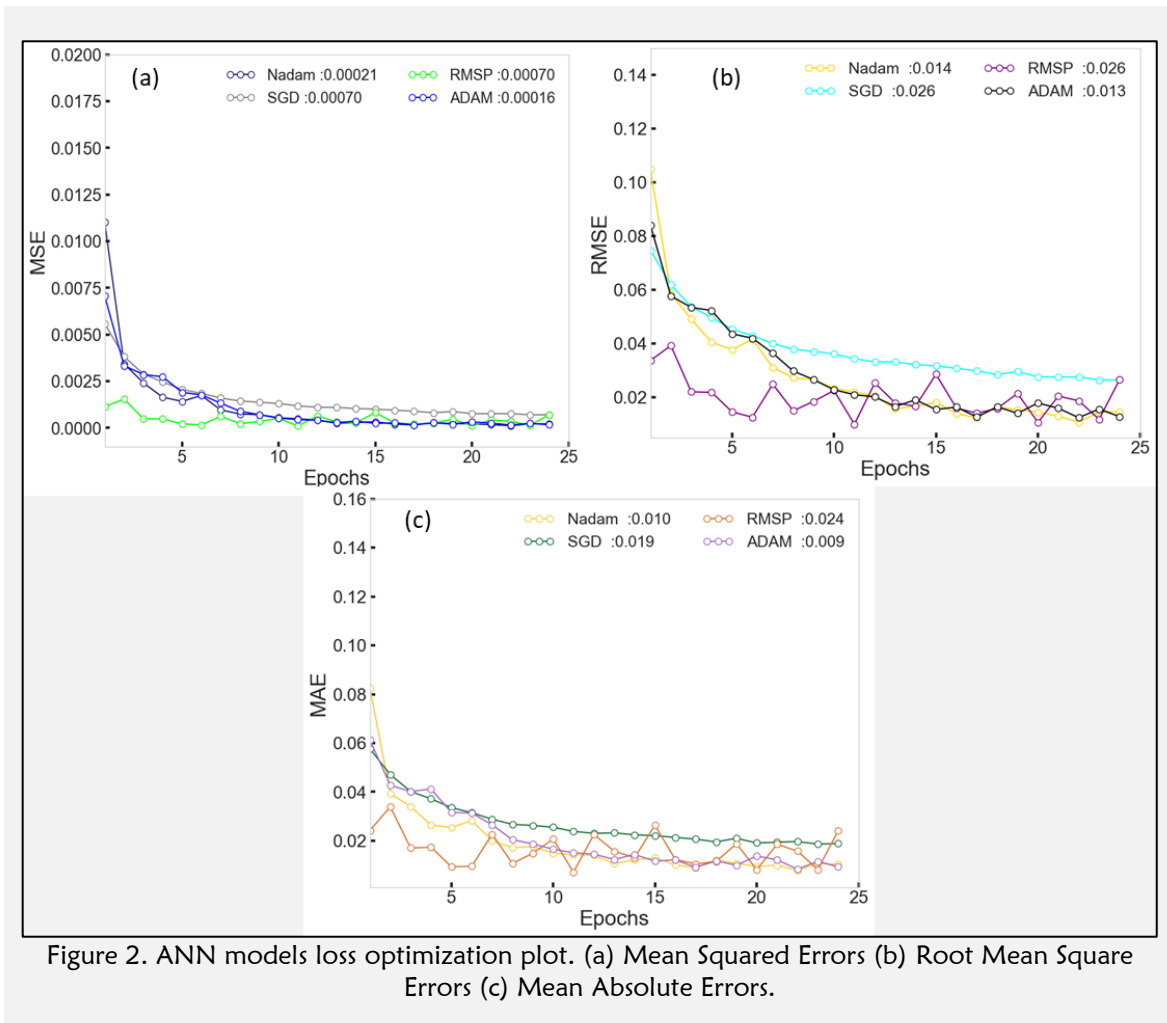
## RESULTS & DISCUSSION

### Model Training

During the optimization process, the performance of the ANN models was iteratively improved through the minimization of loss metrics, specifically Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These metrics were monitored and plotted across 25 training epochs for each of the four optimizers used: NADAM, RMSP, ADAM, and SGD (Figure 2).

The training results demonstrated significant differences in performance among the algorithms. NADAM emerged as the best-performing optimizer, achieving the lowest loss metrics with MSE, RMSE, and MAE values of 0.00021, 0.014, and 0.010, respectively. The consistency in the reduction of these

errors indicates the strong convergence and stability of the NADAM optimizer during the training process.



Following NADAM, the RMSP algorithm showed reasonable performance, with final loss metrics recorded at MSE, RMSE, and MAE values of 0.00070, 0.026, and 0.024, respectively. While RMSP was effective, it lagged behind NADAM in terms of accuracy and error minimization, suggesting that it may be less well-suited for the specific characteristics of this dataset.

The ADAM optimizer performed closely to NADAM, with slightly lower loss metrics: MSE, RMSE, and MAE values of 0.00016, 0.013, and 0.009, respectively. The marginal difference between ADAM and NADAM suggests that both optimizers are well-suited for this type of data, although NADAM's slight edge may be attributed to its adaptive learning rate mechanism, which allows for more precise adjustments during training.

In contrast, the SGD algorithm was the least effective in this context, with MSE, RMSE, and MAE values of 0.00070, 0.026, and 0.019, respectively. Despite being a simpler and less computationally intensive optimizer, SGD struggled with the complex relationships inherent in the data, as evidenced by its higher error rates.

Table 3. Loss feedback after 25 training epochs of ADAM, NADAM, RMSP, and SGD algorithms

	MSE	RMSE	MAE
NADAM	0.00021	0.014	0.01
RMSP	0.0007	0.026	0.024
ADAM	0.00016	0.013	0.009
SGD	0.0007	0.026	0.019



These results are summarized in Table 3, providing a clear comparison of the performance metrics across the four algorithms after 25 training epochs. The findings indicate that the NADAM and ADAM optimizers are particularly effective for this application, given their ability to minimize error and improve model accuracy. This aligns with existing research that highlights the strengths of these adaptive algorithms in handling complex and non-linear data patterns, such as those encountered in well log analysis.

### Test in TMG-04 and performance metrics

The performance of the developed models was further evaluated by applying them to the test well TMG-04, where the predicted density logs were compared with the actual density logs. As illustrated in Figure 3, the results of the five density log predictions, performed using the developed models on well TMG-04, were superimposed over the actual density log to visualize the accuracy of the predictions. The comparison between the predicted logs and the measured logs highlights the effectiveness of the models, as well as their respective strengths and weaknesses.

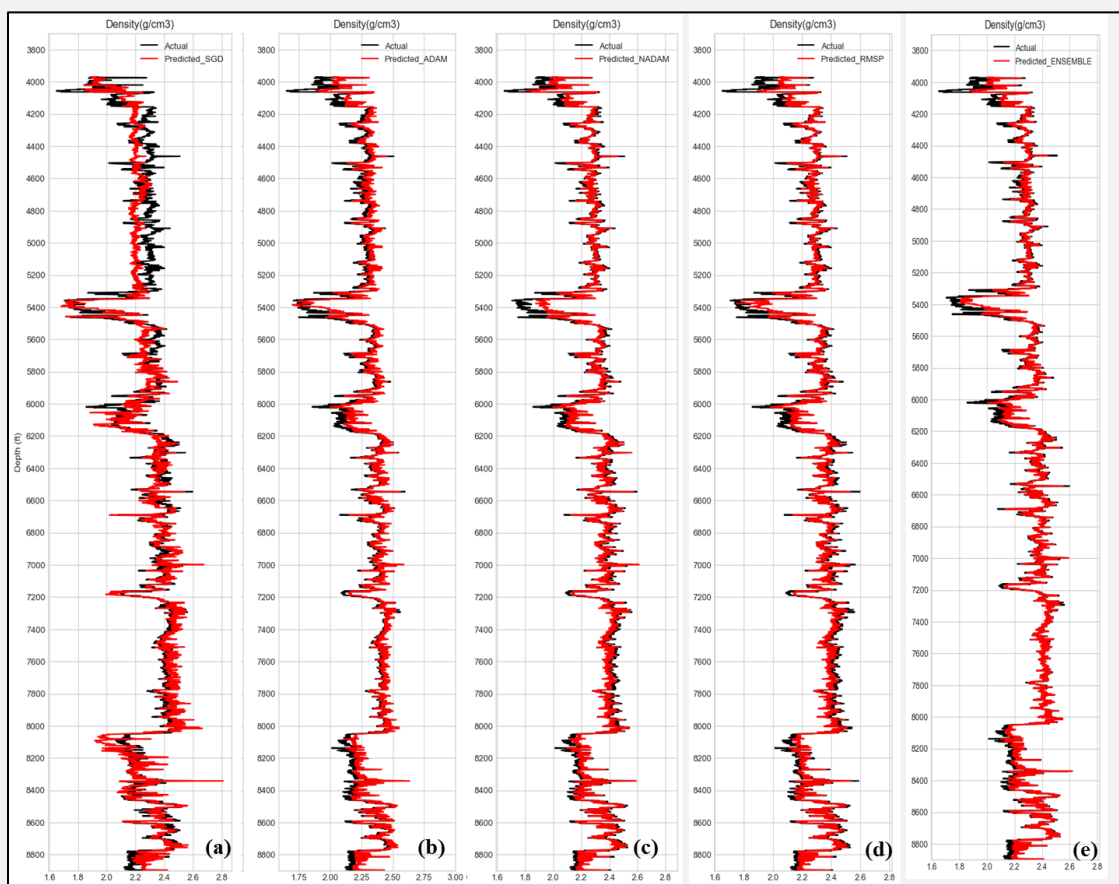


Figure 3. Predicted RHOB log (red line) superimposed on actual RHOB log (black line) in TMG-04 used for test: (a) SGD (b) ADAM (c) NADAM (d) RMSP (e) Ensemble.

To quantitatively assess the accuracy of the predictions, four key performance metrics— $R^2$ , MAE, MSE, and RMSE—were used to evaluate each model. These metrics provided insight into the degree to which each model's prediction deviated from the actual data, thus allowing for a clear comparison of model performance.

The results demonstrated that the Ensemble model outperformed the individual models, showcasing its robustness in accurately predicting the density log. This model was developed by averaging the predictions of the four individual models with specific weights assigned based on their performance: 0.17 for NADAM, 0.28 for RMSP, 0.05 for ADAM, and 0.50 for SGD. The Ensemble model achieved an  $R^2$  of 0.9051, the highest among all models, along with the lowest MAE of 0.0248, MSE of 0.0017, and RMSE of 0.0413.

In comparison, the individual models—NADAM, RMSP, ADAM, and SGD—showed varying levels of accuracy. The NADAM model achieved an  $R^2$  of 0.8832, MAE of 0.0262, MSE of 0.0021, and RMSE of 0.0458. The RMSP model had similar performance, with an  $R^2$  of 0.8858, MAE of 0.0309, MSE of 0.0020, and RMSE of 0.0453. The ADAM model, although close in performance to NADAM and RMSP, was slightly less accurate, with an  $R^2$  of 0.8747, MAE of 0.0278, MSE of 0.0022, and RMSE of 0.0474. The SGD model performed comparably well, achieving an  $R^2$  of 0.8986, MAE of 0.0270, MSE of 0.0018, and RMSE of 0.0427.

The performance metrics are summarized in Table 4, which underscores the superiority of the Ensemble model over the individual models. The Ensemble's ability to combine the strengths of each model through weighted averaging allowed it to produce a more accurate and reliable prediction of the density log for well TMG-04, making it the preferred choice for this application.

Table 4. Performance metrics

Models	R2	MAE	MSE	RMSE
Ensemble	0.9051	0.0248	0.0017	0.0413
NADAM	0.8832	0.0262	0.0021	0.0458
RMSP	0.8858	0.0309	0.0020	0.0453
ADAM	0.8747	0.0278	0.0022	0.0474
SGD	0.8986	0.0270	0.0018	0.0427

### Final Prediction of RHOB in TMG-09 Well

The Ensemble model was employed to predict the density log (RHOB) in the TMG-09 well, and the results are depicted in Figure 4. The mean absolute error (MAE) of our model was calculated to be 0.0248  $\text{g/cm}^3$ . Given that the density of the reservoir sand ranges from 2.0 to 2.6  $\text{g/cm}^3$ , the uncertainty of our model does not exceed 1.3%. This low uncertainty level demonstrates that the Ensemble model is highly effective for predicting the density log in the Niger Delta or in other fields with similar geological settings, provided that the model is supplied with appropriate input parameters such as DT, SW, Depth, and GR.

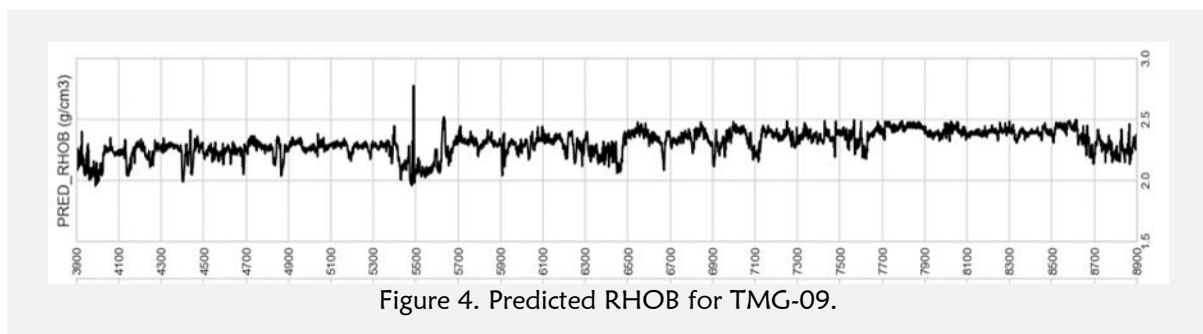


Figure 4. Predicted RHOB for TMG-09.

The findings from this study are consistent with those of Saikia & Baruah (2023), who introduced an ensemble modeling solution named "SEMoRC." This approach effectively predicted various petrophysical properties using seismic data, outperforming conventional models that struggled to achieve the same results. Their model, applied to a real-world dataset, demonstrated superior performance in terms of RMSE and correlation coefficient when compared to its constituent machine learning models.

In the current study, while the NADAM optimizer exhibited the best individual performance, the Ensemble model ultimately outperformed all the individual models. This superiority is attributed to the synergistic effect of combining multiple optimization techniques, which enables the model to capture diverse aspects of the data and, consequently, enhance overall performance.

### CONCLUSION

This study focused on generating a density log (RHOB) using machine learning to enhance reservoir characterization in a gas field within the Niger Basin, utilizing data from five available wells. The depth and three lithologic and petrophysical logs—namely, Gamma Ray (GR), Sonic Transit Time (DT), and Water Saturation (SW)—were employed as predictors to synthesize the RHOB in the TMG-09 well, where it was missing. Four models were trained and validated, and their performance was tested on the TMG-04 well, which had measured RHOB data to evaluate how closely the predicted values

aligned with the actual data.

The test results revealed that the mean absolute errors (MAE) of the predicted density at one-foot intervals in the TMG-04 well were 0.0262, 0.0309, 0.0278, and 0.0270 for the models built using the NADAM, RMSP, ADAM, and SGD optimizers, respectively. The corresponding  $R^2$  values between the synthesized and actual data were 0.8832, 0.8858, 0.8747, and 0.8986. For the final prediction, an Ensemble model was created by combining these individual models into a weighted average, with weights of 0.17 for NADAM, 0.28 for RMSP, 0.05 for ADAM, and 0.50 for SGD. The Ensemble model yielded a MAE of 0.0248, which was lower than that of any individual algorithm, and the highest  $R^2$  value of 0.9051.

The results indicate that the correlation analysis demonstrated that GR, DT, SW, and depth had a significant influence on the target parameter (RHOB), making them suitable input variables. Including depth as an input in the Artificial Neural Network (ANN) improved the model's performance, with the coefficient of determination ( $R^2$ ) increasing from 0.8751 to 0.9051 and the MAE decreasing from 0.0261 to 0.0248 when comparing synthesized and actual data. The Ensemble model outperformed the individual models based on NADAM, RMSP, ADAM, and SGD optimizers. Even with limited data, the Ensemble model offers valuable insights for exploration and production companies, aiding in informed decision-making. The Ensemble model can effectively predict the density log in the Niger Delta or other fields with similar geological settings when supplied with DT, SW, Depth, and GR. While the study examined data from five wells, additional data are needed to further evaluate and enhance the model's accuracy and reliability.

This conclusion highlights the key findings and implications of the study, emphasizing the potential of the Ensemble model for practical applications in reservoir characterization.

## ACKNOWLEDGEMENTS

This study was funded by the Pan African University Institute of Life and Earth Sciences (including Health and Agriculture). We extend our sincere gratitude for their support and contribution to our research.

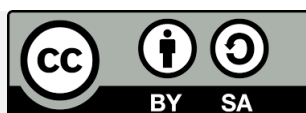
## Declarations

Competing Interests: On behalf of all the co-authors, the corresponding author hereby states that there is no conflict of interest.

## REFERENCES

- Abuh, F. A., Akpabio, J. U., & Okon, A. N. (2023). Machine Learning-Based Models for Basic Sediment & Water and Sand-Cut Prediction in Matured Niger Delta Fields. *Journal of Energy Research and Reviews*, 15(2), 70-93. <https://doi.org/10.9734/jenrr/2023/v15i2310>
- Adegoke, O. S., Oyebamiji, A. S., Edet, J. J., Osterloff, P. L., & Ulu, O. K. (2017). *Chapter 2 - Geology of the Niger Delta Basin* (pp. 25–66). Elsevier. <https://doi.org/10.1016/B978-0-12-812161-0.00002-8>
- Adhari, M. R., & Kardawi, M. Y. (2022). Estimation of density log and sonic log using artificial intelligence: an example from the Perth Basin, Australia. *Journal of Geoscience, Engineering, Environment, and Technology*, 7(4), 158-166. <https://doi.org/10.25299/jgeet.2022.7.4.10050>
- Aggrey, G. H., & Davies, D. R. (2007, September). Tracking the state and diagnosing downhole permanent sensors in intelligent-well completions with artificial neural network. In *SPE Offshore Europe Conference and Exhibition* (pp. SPE-107198). SPE. <https://doi.org/10.2118/107198-MS>
- Ahmadi, M., & Chen, Z. (2020). Machine learning-based models for predicting permeability impairment due to scale deposition. *Journal of Petroleum Exploration and Production Technology*, 10, 2873-2884. <https://doi.org/10.1007/s13202-020-00941-1>
- Al-Bulushi, N., King, P. R., Blunt, M. J., & Kraaijveld, M. (2009). Development of artificial neural network models for predicting water saturation and fluid distribution. *Journal of Petroleum Science and Engineering*, 68(3-4), 197-208. <https://doi.org/10.1016/j.petrol.2009.06.017>
- Chen, H. H., Huang, B., Liu, F., & Chen, W. G. (2017). Principles and applications of machine learning. *Chengdu: University of Electronic Science and Technology Pres*, 2-19.
- Fagbemi, O. I., Olayinka, A. I., Oladunjoye, M. A., & Edigbue, P. I. (2024). Focused reservoir characterization: analysis of selected sand units using well log and 3-D seismic data in Kukih'field, Onshore Niger Delta, Nigeria. *Scientific Reports*, 14(1), 13763. <https://doi.org/10.1038/s41598-024-56100-7>
- Kim, S., Kim, K. H., Min, B., Lim, J., & Lee, K. (2020). Generation of synthetic density log data using deep learning algorithm at the Golden field in Alberta, Canada. *Geofluids*, 2020(1), 5387183. <https://doi.org/10.1155/2020/5387183>
- Lim, J-S. (2005). Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea. *Journal of Petroleum Science and Engineering*, 49(3-4), 182-192.

- <https://doi.org/10.1016/j.petrol.2005.05.005>
- Liu, J.-J. & Liu, J.-C. (2021). An intelligent approach for reservoir quality evaluation in tight sandstone reservoir using gradient boosting decision tree algorithm - A case study of the Yanchang Formation, mid-eastern Ordos Basin, China. *Marine and Petroleum Geology*, 126, 104939. <https://doi.org/10.1016/j.marpetgeo.2021.104939>
- Long, W., Chai, D., & Aminzadeh, F. (2016, May). Pseudo density log generation using artificial neural network. In *SPE Western Regional Meeting* (pp. SPE-180439). SPE. <https://doi.org/10.2118/180439-MS>
- Maju-Oyovwikowhe, E. G., & Osayande, A. D. (2023). Hydrocarbon evaluation and distribution in Well-X and Well-Y in the Niger Delta Basin: Findings and validation through porosity comparison. *Scientia Africana*, 22(1), 255-278. <https://doi.org/10.4314/sa.v22i1.22>
- Miri, R., Sampaio, J., Afshar, M., & Lourenco, A. (2007, June). Development of artificial neural networks to predict differential pipe sticking in Iranian offshore oil fields. In *SPE International Oil Conference and Exhibition in Mexico* (pp. SPE-108500). SPE. <https://doi.org/10.2118/108500-MS>
- Mohaghegh, S. (2000). Virtual-intelligence applications in petroleum engineering: Part 1—Artificial neural networks. *Journal of Petroleum Technology*, 52(09), 64-73. <https://doi.org/10.2118/58046-JPT>
- Ogbamikhumi, A., & Omorogieva, O. M. (2021). Rock property modelling and sensitivity analysis for hydrocarbon exploration in OSSY field, Niger Delta Basin. *Journal of Petroleum Exploration and Production*, 11(4), 1809-1822. <https://doi.org/10.1007/s13202-021-01130-4>
- Origbo, O. A., & Mbachu I. I. (2024). Forecasting Dead Oil Viscosity Using Machine Learning Processes for Niger Delta Region, *International Journal of Current Science Research and Review*, 7(1), 820-830. <https://doi.org/10.47191/ijcsrr/V7-i1-81>
- Ozbayoglu, E. M., & Ozbayoglu, M. A. (2007). Flow pattern and frictional-pressure-loss estimation using neural networks for UBD operations. In *SPE/IADC Managed Pressure Drilling and Underbalanced Operations Conference and Exhibition* (pp. SPE-108340). SPE. <https://doi.org/10.2118/108340-MS>
- Rajabi, M., Beheshtian, S., Davoodi, S., Ghorbani, H., Mohamadian, N., Radwan, A. E., & Alvar, M. A. (2021). Novel hybrid machine learning optimizer algorithms to prediction of fracture density by petrophysical data. *Journal of Petroleum Exploration and Production Technology*, 11(12), 4375-4397. <https://doi.org/10.1007/s13202-021-01321-z>
- Saeedi, A., Camarda, K. V., & Liang, J. T. (2007). Using neural networks for candidate selection and well performance prediction in water-shutoff treatments using polymer gels—a field-case study. *SPE Production & Operations*, 22(04), 417-424. <https://doi.org/10.2118/101028-PA>
- Saikia, P., & Baruah, R. D. (2023). Stacked ensemble model for reservoir characterisation to predict log properties from seismic signals. *Computational Geosciences*, 27(6), 1067-1086. <https://doi.org/10.1007/s10596-023-10248-9>
- Saikia, P., Baruah, R. D., Singh, S. K., & Chaudhuri, P. K. (2020). Artificial Neural Networks in the domain of reservoir characterization: A review from shallow to deep models. *Computers & Geosciences*, 135, 104357. <https://doi.org/10.1016/j.cageo.2019.104357>
- Saporetti, C. M., Goliatt, L., & Pereira, E. (2021). Neural network boosted with differential evolution for lithology identification based on well logs information. *Earth Science Informatics*, 14(1), 133-140. <https://doi.org/10.1007/s12145-020-00533-x>
- Tamhane, D., Wong, P. M., Aminzadeh, F., & Nikravesh, M. (2000, April). Soft computing for intelligent reservoir characterization. In *SPE Asia Pacific conference on integrated modelling for asset management* (pp. SPE-59397). SPE. <https://doi.org/10.2118/59397-MS>
- Tugwell, K. W., & Livinus, A. (2023). Predictive models for oil in place for oil rim reservoirs in the Niger Delta using machine learning approach. *Petroleum & Petrochemical Engineering Journal*, 7(3), 000361. <https://doi.org/10.23880/ppej-16000361>
- Tuttle, M. L., Charpentier, R. R., & Brownfield, M. E. (1999). *The Niger Delta Petroleum System: Niger Delta Province, Nigeria, Cameroon, and Equatorial Guinea, Africa* (pp. 99-50). US Department of the Interior, US Geological Survey. <https://doi.org/10.3133/ofr9950H>



Copyright (c) 2024 by the authors. This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).