Development of empirical correlation for predicting pore pressure from well log data through multiple linear regression analysis

Abstract

Accurate pore pressure estimation is essential for safe drilling practices and effective reservoir modeling, especially in regions where overpressure from disequilibrium compaction poses significant challenges. These challenges can lead to issues such as fluid loss, kicks, differential pipe sticking, heaving shale, and blowouts. Traditional methods often fall short in capturing the complex relationships between formation parameters and pore pressure. This study utilizes a machine learning (ML) approach to capture these intricate relationships, developing two empirical correlations for pore pressure prediction. The first correlation includes lithological information (sand and shale), and the second does not. Both correlations are derived from a linear regression model fitted to the well log datasets. This study utilized data collected from three wells in the Northern Carnarvon Basin of Australia. It includes parameters such as sonic interval transit time, density, gamma-ray, depth, and well diameter. Wells 1 and 2 contributed approximately 22,038 data points, which were divided into 85% for model training and 15% for validation. The 8,860 data points from Well 3 were utilized to evaluate the model's accuracy.

The results, assessed using statistical metrics, including root mean square error (RMSE), mean absolute relative error (MARE), and mean relative error (MRE), show that the newly developed correlations perform better than existing ones. The first model, which includes lithological data, demonstrated promising accuracy with an RMSE of 352.208. The second model, which does not include lithological data, surpassed the first with an RMSE of 342.105. These developed correlations offer enhanced predictive capabilities and effectiveness, making them suitable for forecasting pore pressure in real-world drilling scenarios.

**Keywords**: Formation pore pressure**,** Abnormalpressure**,** Multiple linear regression, Reservoir modelling, Machine learning

# INTRODUCTION

Pore pressure refers to the pressure exerted by fluids within the pore spaces of rock formations, also called formation pressure. Depending on its magnitude, pore pressure can be classified as either normal or abnormal. Normal pore pressure ranges from 0.433 psi/ft in freshwater to 0.465 psi/ft in saltwater, while abnormal pore pressure, due to geological and hydrodynamic factors, can be over pressured (higher) or under pressured (lower) than normal pressure (Swarbrick et al., 1998). Overpressure in oil and gas drilling can arise from various mechanisms including compaction disequilibrium, chemical diagenesis, differential density effect, and fluid expansion which poses significant risks, including kicks, blowouts, lost circulation, and mud loss. As these events can lead to increase non-productive time, hazardous accidents, and increased project costs. Thus, precise pore pressure prediction is essential for safe and cost-effective drilling operations, reducing risks and enhancing project efficiency.

Abnormal formation pressure can be detected and estimated using predictive (pre-drill) methods like correlations from nearby wells and seismic data analysis. During drilling, monitoring drilling parameters and using measurement tools can provide real-time indications. Post-drill, wireline logging and formation testing can verify the accuracy of pore pressure estimates.

Many researchers have carried out studies on identifying abnormal pore pressures. The focus on this issue highlights the significance of the information and the challenges faced in developing a method to accurately deliver this information when it is most urgently required. Hottmann and Johnson (1965) first predicted pore pressure from shale properties using well log data. This technique identifies deviations from normal compaction trends as abnormal pressure. Afterwards, other researchers presented empirical equations using sonic transit time, porosity resistivity, and other well log data for pore pressure prediction. Ham (1966) equivalent depth method identifies abnormal pressure zones by comparing it to normal pressure in a nearby formation. Eaton (1975) proposed two empirical equations to quantify the pore pressure using well log data from sonic log and resistivity log presented in equation 1.0 and 2.0

.1.0

2.0

where is the pore pressure gradient (psi/ft); OBG is the overburden gradient (psi/ft);  is the hydrostatic gradient (psi/ft);  is the normal compaction trend line (s/ft); is the observed sonic log value (s/ft); X is the exponent value which is dependent on formation properties). is the resistivity value obtained from logs; is resistivity of normal compaction trend line.

Bowers (1995) established a power relationship between sonic velocity and effective stress in petroleum basins, employing well logging data to determine vertical effective stress and pore pressure. The methods developed by Eaton and Bowers are among the most commonly used conventional approaches for predicting pore pressure, emphasizing compaction disequilibrium as the primary mechanism behind the generation of overpressure.

Over the past few decades, several data-driven methodologies were developed to estimate subsurface pressure from well log data, seismic data, and drilling parameters. The utilization of the following Machine learning techniques for predicting pore-pressure often fall short of providing an empirical correlation for real-time practical application, unless integrated into software, highlighting the drawbacks of black-box ML models. Hao Yu et al. (2020) developed a machine learning method for pore-pressure prediction from well logs, using a nonparametric multivariate model of petrophysical properties, a year after, Abdelaal et al. (2021) developed three predictive models for real-time pore pressure gradient prediction from mechanical and hydraulic drilling parameters, using support vector machines (SVM), functional networks, and random forest (RF). AE Radwan et al**.** (2022) employed machine learning techniques to predict pore pressure from geophysical logs in Mangahewa gas field, New Zealand using decision tree, Adaboost (ADA), random forest (RF), and transparent open box (TOB).

Huayang Li et al. (2023) also employed machine learning to predict pore pressure in high-pressure reservoir zones, achieving 95% accuracy using KNN, Extra Trees, Random Forest, and LightGBM algorithms. Deng S. et al. (2024) improved pore pressure prediction accuracy using machine learning and optimization algorithms, integrating IGWO-MLP model for superior performance and high R-squared values.

**2.0 METHODOLOGY**

**2.1 DATA DESCRIPTION AND ANALYSIS**

The datasets utilized in this study comprise well logging data from three wells located in the Northern Carnarvon Basin, Australia. The three wells data consisting of 30,898 data points from the study area were analyzed and used to develop the workflow presented in this paper. With well 1 contributing 11,876 data points with a depth range of 457.6572  4,016.1972m, well 2 contributed 10,162 data points with a depth range of 690.8292  4,657.9536m and well 3 contributed 8,860 data points with a depth range of 805.434  4,104.894m. The dataset includes features such as sonic time difference, density, gamma-ray, depth, and well diameter, which are intrinsically linked to the physical properties of the rock layer. The dataset (well 1 and 2) with data points of 22038 is split into an 85% training set and a 15% validation set, with well 3 serving as the blind testing dataset for independent evaluation of the model's performance. Table 1.0 and 2.0 shows a statistical summary of the training and testing dataset. Fig 1.0 shows a scatter pair plots which help to visualize the distribution of variables and identify potential trends.

**Table 1.0** Shows A Statistical Summary of The Training Datasets Features

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DEPTH. Ft | CALI. In | DT.US/F | GR.GAPI | RHOB.G/C3 | PP.psi |
| Count | 22038.0000 | 22038.0000 | 22038.0000 | 22038.0000 | 22038.0000 | 22038.0000 |
| Mean | 9806.456534 | 11.877427 | 86.504113 | 83.767319 | 2.486612 | 5101.192241 |
| Std | 2406.636361 | 1.883954 | 11.414946 | 31.265961 | 0.137513 | 1121.135852 |
| Min | 1501.500048 | 7.663616 | 51.351720 | 40.000000 | 1.568559 | 838.865505 |
| Max | 15282.0008 | 22.640000 | 134.619900 | 234.330000 | 3.160000 | 9472.135973 |

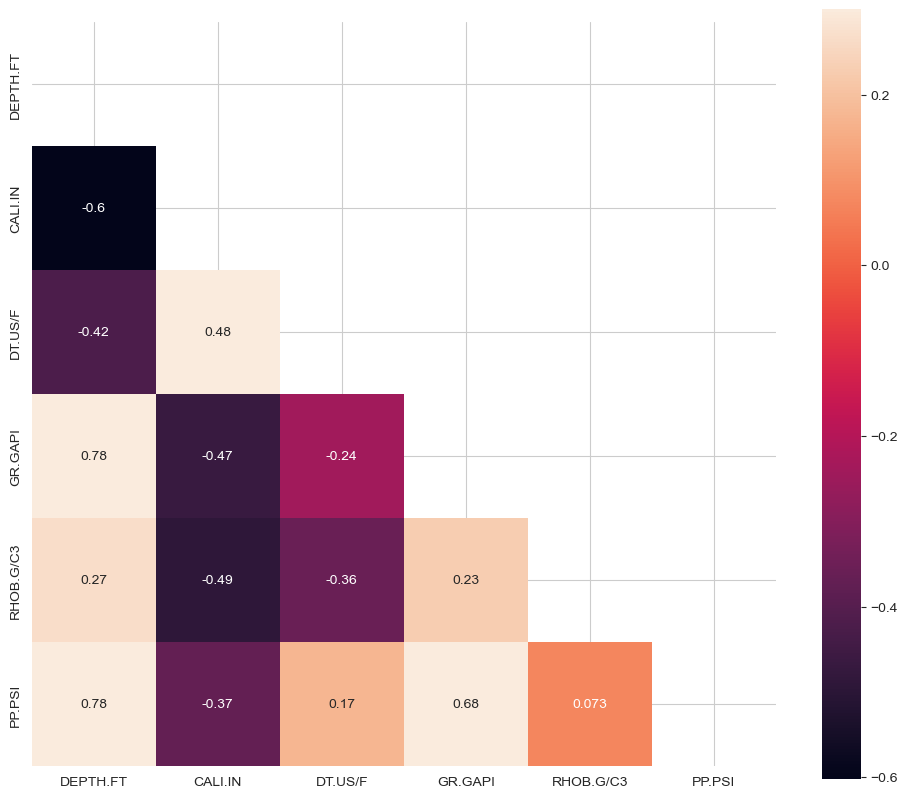
**Table 2.0** Shows A Statistical Summary of The Testing Datasets Features

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DEPTH. ft | CALI. In | DT.US/F | GR.GAPI | RHOB.G/C3 | PP.psi |
| Count | 8860.0 | 8860.0 | 8860.0 | 8860.0 | 8860.0 | 8860.0 |
| Mean | 10569.1603 | 13.899537 | 89.653105 | 8860.0 | 2.486048 | 5933.361139 |
| Std | 1639.4371 | 2.216614 | 12.189492 | 34.602248 | 0.129586 | 843.429689 |
| Min | 2642.5000 | 8.735000 | 46.180000 | 40.000000 | 1.320000 | 1308.379001 |
| Max | 13467.5004 | 24.780000 | 133.350000 | 208.380000 | 3.336000 | 8981.648788 |

****

**Fig.1.0** A Scatter pair-plot of the datasets utilized in this study

This study uses univariate and model-based feature selection to predict pore pressure. Pearson correlations in Fig. 2.0 show depth decreases pore pressure, larger caliper readings and gamma-ray values indicate lower and higher pressures, and higher sonic travel times slightly increase pressure. Two predictive models were developed, one including all relevant features and the other excluding gamma-ray.



**Fig. 2.0** Pearson correlation of the pore pressure features

**2.2 MATHEMATICAL CORRELATION MODELING**

Two new regression-based correlation models will be developed to compute pore pressure. The first model uses depth, well diameter, Interval Transit Time, gamma-ray, and bulk density as predictors, as shown in equation 3.0. The second model excludes gamma-ray, using depth, well diameter, Interval Transit Time, and bulk density, as described in equation 6.0. Both models aim to accurately predict pore pressure through these variables.

**2.2.1 First Correlation**



 ​……………………….3.0

Where ,  = Formation pore pressure, psi;  = Depth, ft;  = Well diameter, Inches;

 = Sonic interval transit time, ;  = Gamma-ray value, GAPI;  = Bulk density, ;  = unknown correlation constants

The nonlinear function was linearized by introducing the logarithm function in equation 3.0:

  + +  + 4.0

The above equation 4.0 corresponds to the multiple regression equation 5.0

y =  + x1 + x2 + x3 + 5.0

where; y = ; x1 = ; x2 = ; x3  = ;  = ;  = ;  = residual error term

**2.2.3 Second Correlation**



 …………………….6.0

Where,  = Formation pore pressure, psi;  = Depth, ft;  = Well diameter, Inches;  = Sonic interval transit time, ;  = Bulk density, ;  = unknown correlation constants

The nonlinear function was linearized by introducing the logarithm function in equation 6.0:

  + +  + 7.0

The above equation 7.0 corresponds to the multiple regression equation 8.0

y =  + x1 + x2 + x3 + …………………………8.0

where; y = ; x1 = ; x2 = ; x3  = ;  = ;  = residual error term

It is therefore objective to determine the optimal values of constants  to that render the residual term e negligible in the developed regression model. To achieve this, the study harnesses the power of Multiple linear regression Machine Learning algorithms to identify the best constants (global minimum) that minimize the residual error term to zero

**2.3 MACHINE LEARNING MODELS UTILIZED**

This study aim to develop a reliable and easy-to-use correlation for predicting pore pressure using linear machine learning models. To achieve this, three linear machine learning models will be explored. This section provides a basic understanding of the mechanisms these models employed, which will influence the results analyzed later in the study.

**2.3.1 Multiple Linear Regression**

Multiple linear regression is a statistical method used to model the relationship between a dependent variable and several independent variables. Its goal is to determine the optimal linear equation for prediction and inference by estimating coefficients that reduce the sum of squared errors between actual and predicted values. The objective function minimizes the L2 norm of the residuals, allowing for the identification of optimal coefficients.

**2.3.2 Regularized Lasso Regression**

Lasso Regression, which stands for Least Absolute Shrinkage and Selection Operator, is a linear regression method that enhances model performance and feature selection by incorporating regularization. It is particularly useful for large predictor sets as it selects relevant features while reducing overfitting. The method introduces an L1 regularization penalty to the standard least squares objective function, limiting coefficient magnitude and encouraging sparse solutions.

**2.3.3 Regularized Ridge Regression**

Ridge Regression is a linear regression technique that improves model stability and addresses multi-collinearity by incorporating an L2 regularization term and a penalty to the standard least squares objective function. This penalty constrains the sum of coefficient squares, promoting smaller values and reducing multi-collinearity impact. The objective function is to find values of beta that minimize this function, balancing data fit and coefficient magnitude, while considering the L2 norm and regularization parameter.

**2.4 MODEL PERFORMANCE METRICS**

In this study, three statistical metrics were used to validate and assess the accuracy of the new correlations against the existing ones.

**2.4.1 Root Mean Square Error**

The Root Mean Square Error (RMSE) was mathematically calculated using the equation 9.0

 9.0

where 𝑛 is the number of observations, ​ represents the observed value, and ​ is the predicted value.

**2.4.2 Mean Absolute Relative Error**

The Mean Absolute Relative Error (MARE) measuring the average percentage difference between predicted and actual values, was determined using the equation 10

 10

where 𝑛 is the number of observations,  represents the observed value, and  is the predicted value.

**2.4.3 Mean Relative Error**

The Mean Relative Error (MRE) measures the average deviation between predicted and actual values, providing insights into predictive model performance by normalizing the difference with actual values magnitude was mathematically calculated using the equation 11

 11

where 𝑛 is the number of observations, ​ represents the observed value, and ​ is the predicted value.

**3.0 Result and Discussion**

One important criterion for choosing a regression model was the need for an interpretable model with accessible coefficients and intercepts for empirical relationships. Three linear machine learning models were trained and validated on 85% and 15% of the training data set. The Simple Multiple linear regression model outperformed the Ridge and Lasso models, with an RMSE of 0.019514 and 0.019375 on the trained and validated set respectively as shown in Table 3.0. The unknown regression constants  to were therefore extracted from the Multiple Linear Regression model which yielded the following correlations given below in equation 12 and 14;

**3.1 First Correlation:**

 **\*** **​………………….…….12**

Where;  = 0.007557138824091286; = 0.99647091; = -0.0768239

 = 1.03034302; = -0.03406814;  = 0.02280915

**…………………...13**

**3.2 Second Correlation**

**………………………………..14**

Where;  = 0.009850685759312815;  = 0.95617249; = -0.07235208;

 = 1.01772954;  = 0.02359403

 **……………...4.4**

**Table 3.0** Models Performance on Pore pressure Training datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | RMSE-Train | RMSE-Val | MAE-Train | MAE-Val |
| Linear Regression | 0.019514 | 0.019375 | 0.407732 | 0.405713 |
| Ridge | 0.022385 | 0.022266 | 0.444818 | 0.450455 |
| Lasso | 0.105746 | 0.099973 | 2.210725 | 2.128521 |

**3.4 MODEL COMPARISON WITH EXISTING CORRELATION**

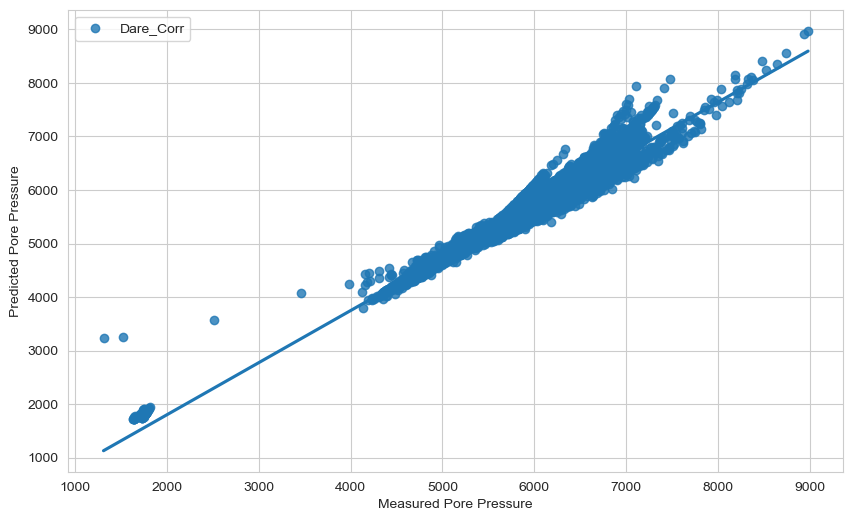
The performance of the newly developed empirical correlations are further compared with the widely used Eaton's correlations carried out on the test datasets, which is crucial to ascertain the extent of improvement achieved by the new correlations in predicting pore pressure using performance metrics.

**3.5 Comparison of the First Correlation with Eatons Correlation On the Test Datasets**

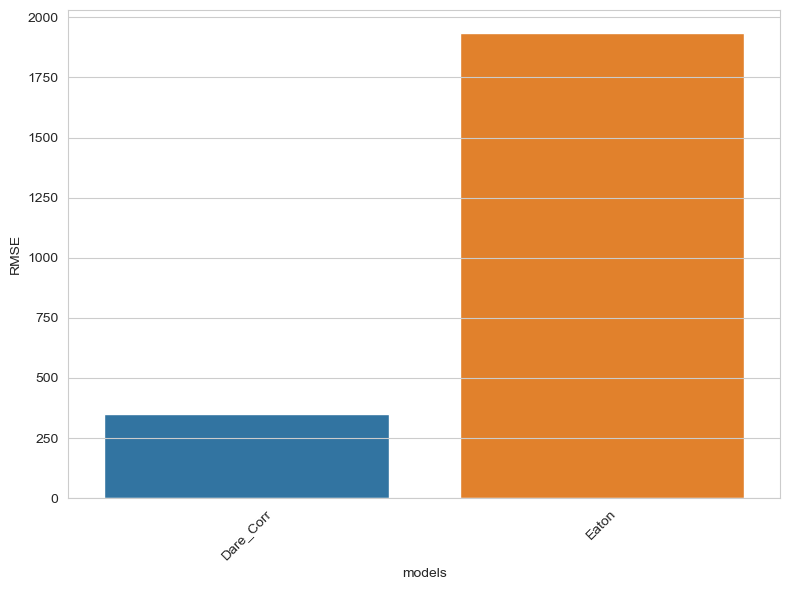
The results in Table 4.0 indicate that the new (first) Correlation model outperforms the Eaton correlation in predicting pore pressure, evidenced by its lower RMSE and MARE values. This new model provides a better fit and more accurate predictions, with a positive MRE indicating overestimation, contrasting with the negative MRE of the Eaton correlation. Figs. 3.0 and 4.0 visually support these findings, showing the new model's superior performance in a scatter plot of measured versus predicted pore pressure and a bar plot comparing RMSE scores, respectively.

**Table 4.0** Showing a comparison of the First correlation developed with Eatons correlation on the Test datasets

|  |  |  |  |
| --- | --- | --- | --- |
|  | RMSE | MARE | MRE |
| New-Correlation | 352.207768 | 5.398975 | 5.120795 |
| Eaton | 1933.965552 | 29.506270 | -28.907812 |

****

**Fig. 3.0**: Cross plot between the measured and the estimated pore pressure on the entire test dataset using the first correlation extracted from the Multiple linear regression model.

****

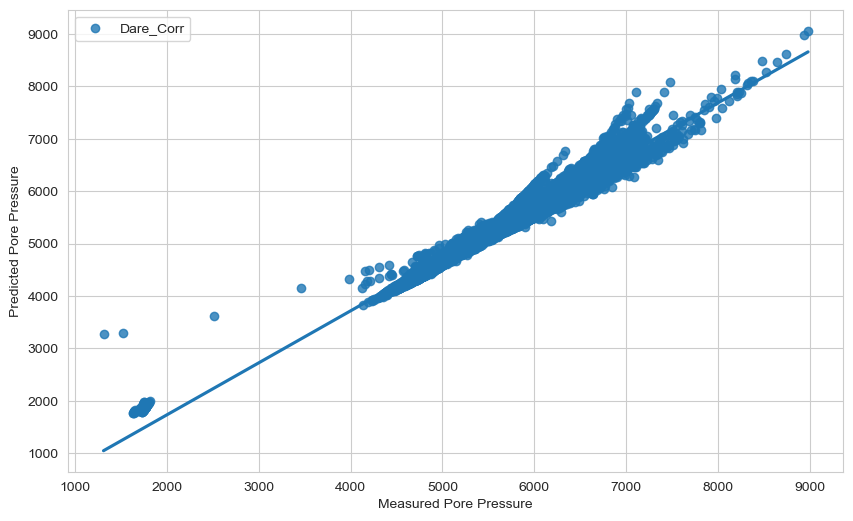
**Fig. 4.0** A bar plot showing the first correlation and Eaton ranking on the testing dataset based on RMSE score

**3.6 Comparison of the Second Correlation with Eatons Correlation On the Test Datasets**

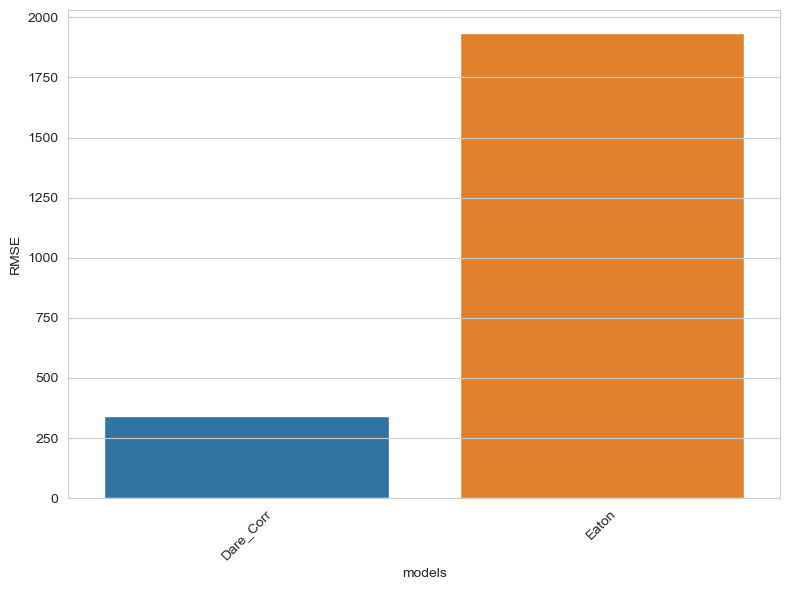
The result obtained from Table 5.0, shows that the New Correlation model outperforms the Eaton correlation in predicting pore pressure, indicated by lower RMSE and MARE values. The New Correlation model offers a better fit and more accurate predictions, with a positive MRE reflecting overestimated values, in contrast to the Eaton correlation's negative MRE, which suggests underestimated pore pressure levels. Figures 5.0 and 6.0 support these findings, depicting the relationships between measured and predicted pore pressures and comparing RMSE scores of the second correlation and Eaton correlation, respectively.

**Table 5.0** Showing a comparison of the Second correlation developed with Eatons correlation on the Test datasets

|  |  |  |  |
| --- | --- | --- | --- |
|  | RMSE | MARE | MRE |
| New-Correlation | 342.104553 | 5.385154 | 5.043188 |
| Eaton | 1933.965552 | 29.506270 | -28.907812 |

****

**Figure 5.0**: Cross plot between the measured and the estimated pore pressure on the entire testing dataset using the second correlation extracted from the Multiple linear regression model.

****

**Fig. 6.0** A bar plot showing the Second correlation and Eaton ranking on the testing dataset based on RMSE score

**3.6 Discussion**

In this study, Eaton's method was used to estimate pore pressure from the acoustic logs of the existing test dataset, assuming a normal compaction trend line derived from the exponential relationship between interval transit time and depth. Overburden pressure was determined from density log measurements, and a mud weight of 9.7ppg was used to determine the hydrostatic gradient. Pore pressure values were computed using Eaton's equation (1) for comparison purposes with an x exponent of 3. The results obtained from both correlations show that the new empirical correlations predict pore pressure more accurately than Eaton's correlation. Their lower RMSE, MARE, and MRE values indicate greater accuracy and reliability, making them better suited for practical applications in regions of overpressure due to disequilibrium compaction in the Northern Carnarvon Basin.

**4.0 Conclusion**

This study utilized machine learning models to develop two empirical correlations for predicting pore pressure using well logging data from three wells in the Northern Carnarvon Basin, Australia. The two correlations were developed from multiple linear regression models by linearizing a nonlinear pore pressure function, and the best coefficient constants were extracted from the linear regression model fitted to the datasets. Both correlations were compared to an existing correlation, and the results showed that the new empirical correlations predict pore pressure more accurately than Eaton's correlation.

The comparative analysis of the two developed correlations revealed that the second correlation was superior, with an RMSE of 342.104553 compared to the first correlation's RMSE of 352.207768, making it the optimal choice. Despite both correlations performing significantly well, the first correlation, which utilized lithology values from gamma-ray logs, is proposed for predicting pore pressure for new wells using existing data. Conversely, the second correlation, effective even without specific lithology information, is proposed as a robust solution for predicting pore pressure in new wells for future drilling operations. This versatility enables decision-makers to tailor their pore pressure prediction methods based on data availability, thereby supporting sound decision-making in oil and gas exploration.

**References**

Abdelaal, A., Elkatatny, S. and Abdulraheem, A., 2021. Data-driven modeling approach for pore pressure gradient prediction while drilling from drilling parameters. *ACS omega*, *6*(21), pp.13807-13816.

Bowers, G. L. (1995). Pore Pressure Estimation From Velocity Data: Accounting for Overpressure Mechanisms Besides Undercompaction. Paper SPE27488, presented at the IADC/SPE Drilling Conference, Dallas, 515530.

Deng, S., Pan, H.Y., Wang, H.G., Xu, S.K., Yan, X.P., Li, C.W., Peng, M.G., Peng, H.P., Shi, L., Cui, M. and Zhao, F., 2024. A hybrid machine learning optimization algorithm for multivariable pore pressure prediction. *Petroleum Science*, *21*(1), pp.535-550.

Eaton, B. A. (1975). The equation for geopressure prediction from well-logs. Paper presented at the fall meeting of the Society of Petroleum Engineers of AIME. 2118/5544-MS

Ham Harold 1966 A method of estimating formation pressures from gulf Coast Well Logs. Gulf Coast Association of Geological Societies Transactions Vol. 16 Pages 185-197

Hottman, C. E., and Johnson, R. K. (1965). "Estimation of Formation Pressures from Log-Derived Shale Properties." Journal of Petroleum Technology, 17(6), 717-722.

Huang, H., Li, J., Yang, H., Wang, B., Gao, R., Luo, M., Li, W., Zhang, G. and Liu, L., 2022. Research on prediction methods of formation pore pressure based on machine learning. *Energy Science & Engineering*, *10*(6), pp.1886-1901.

Radwan, A.E., Wood, D.A. and Radwan, A.A., 2022. Machine learning and data-driven prediction of pore pressure from geophysical logs: A case study for the Mangahewa gas field, New Zealand. *Journal of Rock Mechanics and Geotechnical Engineering*, *14*(6), pp.1799-1809.

Swarbrick R.E. and M.J.Osborne, 1998, Mechanisms that generate abnormal pressure: An overview, in B.E. Law, G.F. Ulmishek and V.I. Slavin, eds, Abnormal pressures in hydrocarbon environments: AAPG Memoir 70, p. 13-34