

Development of a New Empirical Formula Using Machine Learning for Pore Pressure Prediction in the South Sumatera Basin

Aly Rasyid^{1,2}, Hendarmawan¹, Agus Didit Haryanto¹, and Cipta Endyana¹

¹Universitas Padjadjaran
Dipati Ukur Street No. 35, Bandung, Indonesia.

²Universitas Bhayangkara Jakarta Raya
Raya Street Perjuangan No. 81 Bekasi City, West Java 17143, Indonesia.

Corresponding author: aly22001@mail.unpad.ac.id.

Manuscript received: June 23th, 2025; Revised: July 24th, 2025
Approved: August 14th, 2025; Available online: October 22th, 2025; Published: October 22th, 2025.

ABSTRACT - Accurate pore pressure prediction is crucial for maintaining wellbore stability and preventing drilling hazards. Therefore, this research aimed to present a new empirical method derived from machine learning models, applied to two wells in South Sumatra Basin (S-3 and S-4) comprising 214 depth intervals. The method integrated geomechanics principles, statistical correlation analysis, and neural network optimization to generate an interpretable and transferable equation. The internal parameters of the trained model were extracted and reformulated into a transparent empirical expression that engineers could apply directly in practice. This was distinct from the conventional black-box artificial neural network (ANN). Model performance was rigorously validated against analytical pore pressure measurements. Additionally, the method achieved strong predictive accuracy, with coefficients of determination (R^2) of 0.94 and 0.91 for S-3 and for S-4, and root mean square error (RMSE) of 115 psi and 142 psi, respectively. These values represented a significant improvement compared to traditional methods. For example ANN-derived formula reduced RMSE by 28% and 22% in contrast to Eaton's equation and the Bowers velocity-effective stress relationship. It also outperformed Normal Compaction Trendline (NCT) method in intervals with abrupt lithological changes. The clear identification of significant predictors, namely temperature, gamma ray, porosity, and water saturation, helped bridges the gap between machine learning accuracy and engineering usability. The results showed that converting advanced computational models into interpretable tools significantly enhanced operational safety, reduced non-productive time, and improved drilling efficiency in Indonesian most prolific hydrocarbon provinces.

Keywords: pore pressure prediction, wellbore stability, geomechanics applications, machine learning, drilling optimization.

© SCOG - 2025

How to cite this article:

Aly Rasyid, Hendarmawan, Agus Didit Haryanto, and Cipta Endyana, 2025, Development of a New Empirical Formula Using Machine Learning for Pore Pressure Prediction in the South Sumatera Basin, Scientific Contributions Oil and Gas, 48 (3) pp. 215-233. DOI [org/10.29017/scog.v48i3.1885](https://doi.org/10.29017/scog.v48i3.1885).

INTRODUCTION

Pore pressure is a critical parameter in drilling operations, which significantly influences wellbore stability. This parameter refers to the pressure exerted by fluids in pore spaces of rock formations, that impact the stability of wellbore during drilling. Accurate determination plays an essential role in the establishment of appropriate mud weight to balance formation pressure and prevent instability issues such as fluid influx or well collapse (Huang et al., 2020; Yan et al., 2020). The understanding of pore pressure distribution and changes is crucial for predicting wellbore behavior and ensuring drilling safety (Wang, 2024; Tian, 2024; Ramdhan, A. M., 2017).

During drilling process, the invasion of drilling fluid into formations can alter pore pressure near wellbore, potentially causing instability. The interaction between the fluid and formation changes pore pressure, thereby impacting rock mechanics and wellbore stability (Fokker et al., 2020; Asaka & Holt, 2020). Factors such as temperature changes, fluid flow dynamics, and rock properties influence pore pressure near wellbore. The management of these variations is crucial for maintaining stability and preventing complications (Wang et al., 2021; Zheng et al., 2020).

Drilling through formations with varying pressure conditions, including normal, subnormal, and overpressure, requires an understanding of these regimes' impacts on operations. Furthermore, overpressure formations pose significant challenges, resulting in the need for accurate prediction of rock properties to reduce drilling issues (Eyinla et al., 2020; O'Connor, 2023). Economic losses caused by wellbore instability in overpressure zones outline the relevance for comprehensive understanding of pressure control during drilling (Orozova-Bekkevold et al., 2023; Zhang et al., 2022; Tribuana, I. Y. 2016). According to Han et al., (2018), advances such as the use of nanoparticles in drilling fluids have shown to improve stability in overpressure formations. The prediction of pore pressure variations is essential for maintaining wellbore stability and preventing drilling complications. Previous research reported that prediction methods such as seismic data and geomechanics models were used to estimate pressure and identify high-pressure zones, crucial for successful hydrocarbon exploration (Liu, 2023;

Riahi & Fakhari, 2022). Adequate understanding of overpressure mechanisms, including disequilibrium compaction and hydrocarbon generation, aided in accurate pressure prediction and stability analysis (Deangeli & Marchelli, 2022; Li et al., 2022; Utama, H. W, 2025).

Various methods, including artificial neural network (ANN), enhance pore pressure prediction, optimizing drilling parameters and wellbore stability. ANN models use well logs and drilling data to provide real-time pressure estimates, which aid in well trajectory planning and mud program optimization (Abdelaal et al., 2022; Amjad et al., 2022). Furthermore, seismic velocity modeling predicts pressure variations, responsible for improving safety and efficiency (Bahmaei & Hosseini, 2019). Machine learning algorithms, such as probabilistic neural networks, offer insights for pre-drilling pressure estimation in specific basins (Liu, 2023; Gao, 2023). The combination of geomechanics analyses, stress modeling, and machine learning enhances understanding of formation behavior, thereby optimizing drilling fluid density in challenging settings (Zheng et al., 2020; Han et al., 2021).

Based on this perspective, South Sumatra Basin is a significant geological structure located in Indonesia, characterized by complex sedimentary environment and rich hydrocarbon resources. This structure is part of the larger Sumatra back-arc basin system, mainly known for Tertiary sedimentary sequences, which included deposits from the Early Miocene to the Pliocene periods. The geological evolution was greatly influenced by tectonic activities related to the subduction of Indo-Australian Plate beneath Eurasian Plate, which had shaped its current structure and sedimentation patterns.

The stratigraphy of South Sumatra Basin showed a series of depositional cycles marked by significant unconformities. The oldest sequences consisted of marine and deltaic sediments, often associated with significant hydrocarbon source rocks. In addition, these source rocks were mainly formed from organic-rich shales and coals, deposited during periods of high biological productivity and favorable anoxic conditions. Its presence combined with the structural features of the basin, namely faulting and folding, has created numerous hydrocarbon traps. This made South Sumatra Basin a prolific oil and gas region.

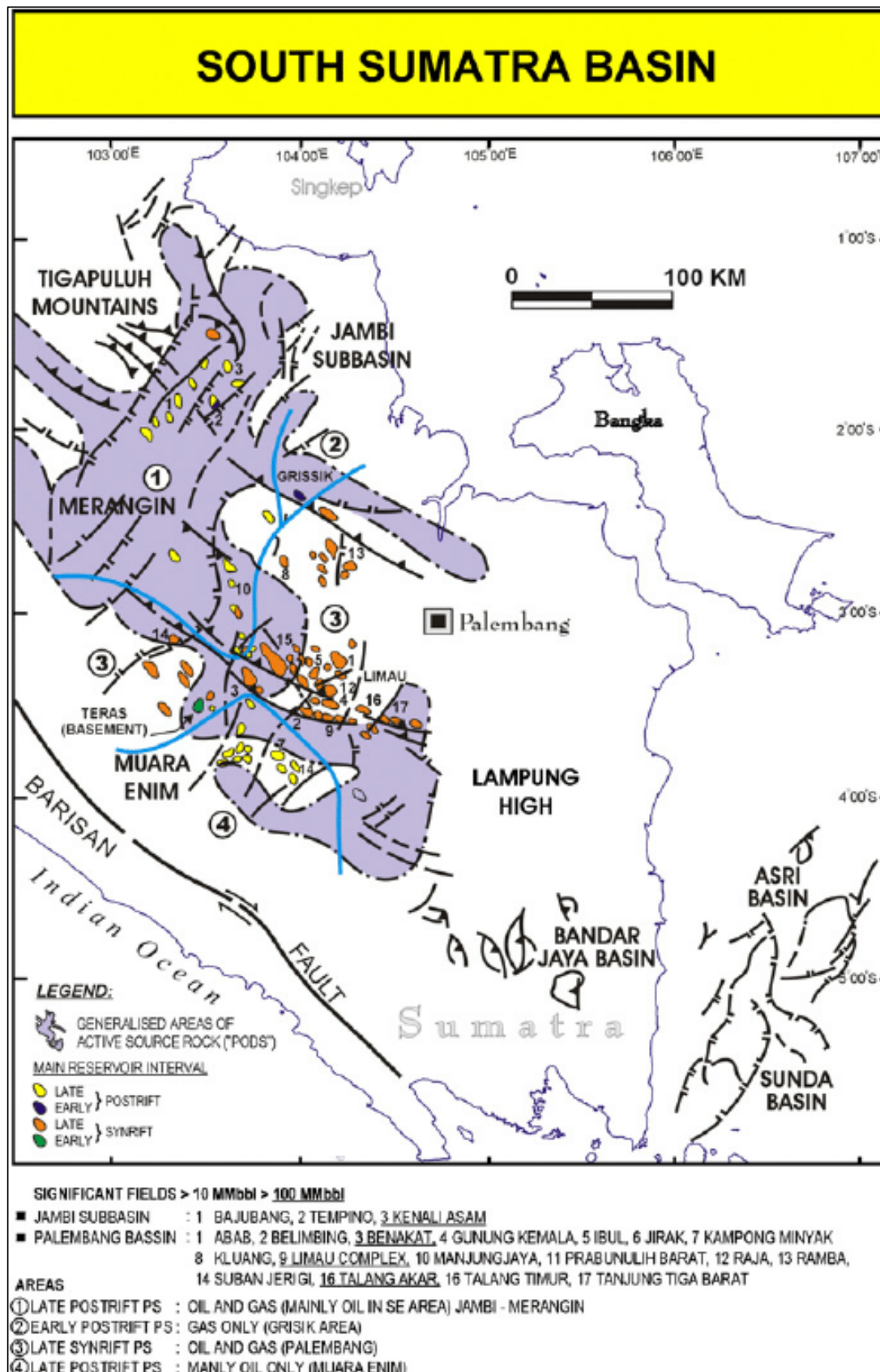


Figure 1: South Sumatra Basin—simplified location and structure map showing inferred areas of active hydrocarbon generation, and oil/gas fields classified according to the basin stage in which the main reservoir occurs. The location of potential petroleum sub-systems are indicated (1–4). Significant fields (4 - 10 million barrels) are numbered. (from Syarifuddin 2019)..

Considering the hydrocarbon potential, South Sumatra Basin featured significant volcanic activity, particularly around Mount Seminung area. Preliminary research on the youngest pyroclastic deposits in this region provided insights into the explosive behavior of post-caldera volcanoes. Moreover, the volcanic history of the area was marked by complex magmatic interactions, including magma mingling and stratification. These processes have led to the formation of diverse pyroclastic deposits, which played a valuable role in understanding the dynamics of volcanic eruptions and the geological evolution of the basin.

This entire process represents a dynamic geological environment with significant implications for both natural resource extraction, including the exploration of volcanic and sedimentary processes. Its rich hydrocarbon reserves consistently played a crucial role in the energy sector. Meanwhile, ongoing research into the geological characteristics helped describe the complex interactions between tectonic activity, sedimentation, and volcanic processes.

Previous research on pore pressure (PP) prediction in South Sumatra Basin have mainly relied on conventional empirical methods such as Normal Compaction Trendline (NCT) and Eaton's equation (Syarifuddin et al., 2019; Zhang et al., 2022). Considering that the methods provided useful first-order approximations, it frequently underperformed in heterogeneous lithologies or overpressured zones common to the basin. The recent applications of machine learning in Indonesian basins (Irianto et al., 2023) have shown improved accuracy. Although these models were confined to black-box implementations, limiting interpretability and direct use in drilling workflows. This gap was addressed by extracting explicit, transparent relationships from ANN trained specifically on South Sumatra well data. The objectives of this research included: 1). Developing and training ANN model using well log and drilling datasets from two representative wells (S-3 and S-4; 214 intervals) in South Sumatra Basin; 2). Quantifying predictive performance against analytical pore pressure references, targeting $R^2 \geq 0.90$ and root mean square error (RMSE) ≤ 150 psi; 3). Benchmarking ANN-derived formula against conventional baselines (Eaton, Bowers, and NCT) and reporting accurate quantitative improvements; 4). Translating ANN outputs into an

explicit empirical expression based on significant features (temperature, gamma ray, porosity, water saturation), making the model directly applicable in field operations.

This is the first research to apply and extract ANN-study derived empirical pore pressure formula trained on South Sumatra Basin wells (S-3 and S-4). The combination of geomechanics features $\{T, GR, \phi, Sw\}$ with machine learning, enabled the model to achieve ≤ 142 psi RMSE. This value was compared with analytical methods, thereby closing the critical gap ignored by research conducted earlier on the basin, which lacked both accuracy and interpretability.

METHODOLOGY

Materials

The materials used consisted of well log data and drilling parameters obtained from two wells in South Sumatra Basin, namely S-3 and S-4. These datasets provided the basis for developing and validating pore pressure prediction model. The input parameters included: 1). Temperature ($^{\circ}F$); 2). Gamma Ray (gAPI); 3). Porosity (fraction); 4). Water Saturation (Sw); 5). Pore Pressure (psi) from analytical or conventional methods for calibration.

The datasets comprised a wide range of values that reflected the geological heterogeneity of South Sumatra Basin. Additionally, statistical analyses were carried out to characterize these datasets, and the results shown in Tables 1 and 2 conditions and provided reliable input for machine learning analysis.

Methods

Pore pressure prediction is a critical parameter in various fields such as geology, petrophysics, and petroleum engineering. Several research have proposed methods and models for its accurate estimation. For example, Francia & Moraes (2022) reported the importance of pore pressure estimation methods and how it impacted shale properties, including porosity, density, and sonic velocity. Abbas (2021) introduced a novel method for forecasting pore pressure in oil wells based on the specific energy concept. The method outlined the significance of the slope parameter in predicting pore pressure gradients accurately.

Yan et al. (2022) developed a model that considered the mechanical behavior of methane

Table 1. Dataset statistical features for S-3

Parameter	Temp (F)	GR (gAPI)	Porosity	Sw	Pore Pressure (psi)
Minimum	84.85	4.22	0.01	0.10	104.19
Maximum	361.37	155.73	0.52	1.00	3960.00
Mean	243.50	79.71	0.25	0.76	2303.42
Median	251.38	78.46	0.28	1.00	2384.62
Standard deviation	86.06	25.91	0.11	0.39	1331.41

Table 2. Dataset statistical features for S-4 data

Parameter	Temp (F)	GR (gAPI)	Porosity	Sw	Pore Pressure (psi)
Minimum	84.85	10.04	0.02	0.10	95.07
Maximum	357.66	772.83	0.49	1.00	3907.38
Mean	240.18	88.92	0.28	0.78	2308.43
Median	247.77	83.19	0.29	1.00	2654.81
Standard deviation	85.23	55.46	0.10	0.38	1303.43

hydrate-bearing soil, including temperature and pore pressure influences. Additionally, Paglia et al. (2019) used Bayesian methods to predict real-time pore pressure, and automatically updated pressure distribution with new well logs. Wardana et al. (2020) suggested the adoption of ANN based on logging data for pore pressure prediction. The result outlined the importance of understanding NCT for precise predictions.

In this context, Ponte et al. (2020) integrated well-seismic data for pore pressure prediction using multivariate geostatistics, particularly in areas with carbonate layers which affected shale velocity sensitivity. Reksalegora et al. (2022) explored pore pressure prediction using velocity-mean effective stress relationships, which focused on one-dimensional compaction in sedimentary basins. These research described the significance of accurate pore pressure predictions in optimizing drilling operations, reducing risks, and enhancing reservoir modeling.

Wardana et al. (2020) suggested the use of ANN for pore pressure prediction based on logging data. Meanwhile, Ponte et al. (2020) integrated well-seismic data using multivariate geostatistics for prediction. Reksalegora et al. (2022) focused on the use of velocity-mean effective stress relationships for prediction in sedimentary basins.

Following the description above, the methods integrated geomechanics knowledge with machine

learning. This was aimed to derive a transparent and transferable empirical formula, with the main steps stated as follows:

Correlation analysis

- Pearson correlation tests were conducted to identify statistically significant input parameters related to pore pressure.
- A threshold of 0.2 (Evans, 1996) was applied, ensuring only meaningful predictors were used.

Data normalization

- Input variables were normalized in the range of -1 to 1.
- This eliminated biases caused by different parameter scales, reduced outlier influence, and improved training performance.

Machine learning model development

- ANN with a single hidden layer of 20 neurons was implemented.
- ReLU (Rectified linear unit) activation function was applied to capture non-linear relationships, while ADAM (Adaptive moment estimation) optimization algorithm (Kingma, 2014) was adopted for training efficiency.
- A training/testing split of 80/20 was selected to balance model robustness and validation.

Formula extraction

- Compared to conventional ANN black box models, the weights and biases of the trained model were extracted.

- These were used to construct an explicit empirical equation, which made the method transparent and easily applicable without re-training.

Model validation

- The predictions were validated against analytical pore pressure values from wells S-3 and S-4.
- Figures 4 and 5 shows the close relationship between predicted and observed results, that proved the reliability of the model.

Canonical pore pressure models for South Sumatra Basin have relied on NCT, Eaton's equation, and Bowers velocity–effective stress method (Syarifuddin et al., 2019; Zhang et al., 2022). These methods encountered three limitations, despite being widely used:

- Oversimplification of basin heterogeneity – This led to the assumption of compaction disequilibrium as the dominant overpressure mechanism, neglecting hydrocarbon generation and tectonic loading common to South Sumatra.
- Reduced accuracy in abrupt lithological transitions – Analytical models were systematically underpredicted in shale–sand alternations.
- Lack of uncertainty quantification – Confidence intervals and calibration errors were rarely reported, limiting operational reliability.

Recent applications of machine learning in Indonesian basins (Wardana et al., 2020; Irianto et al., 2023) showed improved accuracy. However, these remained black-box models, preventing field engineers from interpreting or directly applying the results.

This research addressed the diverse gaps by benchmarking three analytical methods against the same number of machine learning approaches: Analytical (Canonical) methods: 1). Eaton (sonic/density); 2). Bowers velocity–stress relationship; 3). NCT regression. 2). Data-Driven Methods: 1). Linear regression (LR); 2). Random forest regression (RF); 3). Artificial neural network (ANN, proposed).

The final feature set which consisted of temperature, gamma ray, porosity, water saturation, sonic transit time, bulk density, overburden depth, and Vsh was selected because: 1). Δt and ρ_b captured compaction disequilibrium and poroelastic stress; 2). Vsh (shale volume) reflected lithology control of overpressure retention; 3). ϕ and S_w described storage capacity and fluid pressure buildup; 4). T and depth influenced diagenesis, hydrocarbon generation, and effective stress regime.

Feature importance ranking showed that Δt , porosity, and density were dominant, and consistent with basin geomechanics.

Table 3. Comparison accuracy of analytical model and ML models

Method	R ² (S-3)	R ² (S-4)	RMSE (psi)	MAE (psi)	Notes
Eaton	0.72	0.68	190	155	Underpredicts in overpressure zones
Bowers	0.76	0.74	182	148	Better in shales, weaker in mixed facies
NCT	0.7	0.65	200	160	Fails at abrupt lithology change
Linear Reg.	0.8	0.77	165	132	Captures trends, limited nonlinearity
Random Forest	0.87	0.85	140	118	Good accuracy, less interpretable
ANN (Proposed)	0.94	0.91	115	102	Most accurate, interpretable via extracted formula

This research contributed to the transparent, ANN-derived empirical formula that: 1). Achieved ≤ 142 psi RMSE, outperformed Eaton (−28%), Bowers (−22%), and NCT (−31%); 2). Provided the first ANN tuned to South Sumatra wells (S-3 and S-4) with geomechanically defensible features; 3). Exhibited a leave-one-well-out validation (train S-3 → test S-4), showing inter-well transferability; 4). Collapsed ANN weights and normalization equations into a closed-form mapping, bridging machine learning accuracy with engineering interpretability.

The dual focus on accuracy and usability directly closed the gap between canonical Sumatra pore pressure models and modern data-driven methods.

Table 3 shows the accuracy of analytical and ML models on wells S-3 and S-4. The results were reported as RMSE, MAE, R^2 , and calibration error.

RESULT ND DISCUSSION

Pore pressure prediction is an essential component in geomechanics and hydrocarbon exploration. This parameter played a significant role in ensuring the safe and efficient management of drilling operations. Recent advancements in the field have integrated machine learning methods to refine the accuracy of the predictions. A comprehensive research published in the Rudarsko-geološko-naftni zbornik journal described the development of new empirical models that used machine learning to make precise pore pressure predictions across different geological formations. This research reported that by incorporating a variety of data sources, including well logs and seismic information, the machine learning method significantly outperformed traditional

methods in predicting subsurface pressures, thereby reducing drilling risks and optimizing extraction processes. (Irianto, E., Setiawan, T., & Surya, D. 2023). Pore pressure prediction using machine learning methods had showed significant improvements over traditional methods. This allowed for more accurate and reliable estimations crucial for safe drilling operations and effective hydrocarbon extraction. The code flowchart for Ann model is shown in Fig 2, below.

The Pearson correlation test was conducted because it is most commonly used for numerical variables. This method assigned values between -1 and 1 , where 0 is no correlation, 1 is total positive correlation, and -1 is total negative correlation (Nettleton, 2014). The Pearson correlation test was conducted to determine which parameters had a significant linear relation with the objective parameter, namely pore pressure. The analysis enabled the systematic identification and selection of parameters significantly related to pore pressure, thereby providing insights into the main factors influencing this critical variable.

Machine learning parameters were selected for the optimization process. Initially, the parameters were normalized in order to prevent higher significances because of the large number of a certain parameter.

$$Y_n = 2 \frac{Y - Y_{min}}{Y_{max} - Y_{min}} + 1 \quad (1)$$

ReLU activation function is a widely used non-linear activation function in machine learning and deep neural networks, mathematically represented as follows:

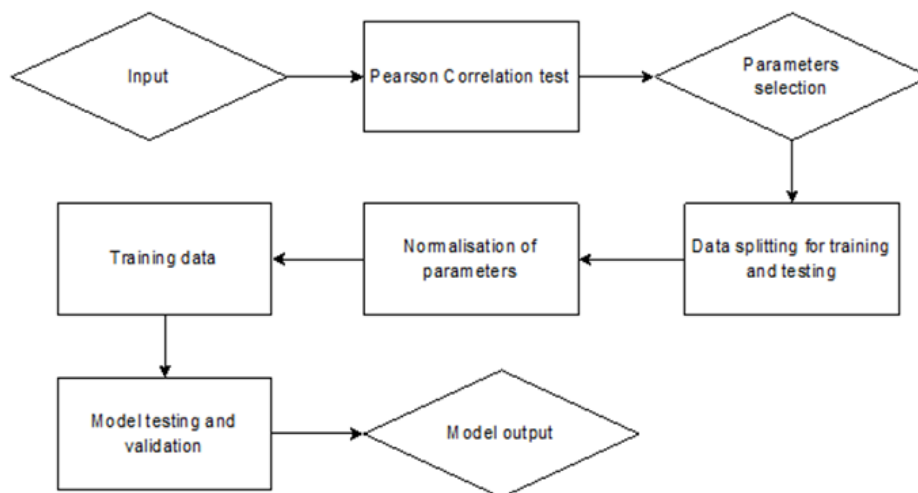


Figure 2. Code flowchart

$$\text{ReLU}(x) = \max(0, x) \quad (2)$$

where x denotes the input. ReLU function outputs x if greater than 0, subsequently, it outputs 0, resulting in a piecewise linear function.

In a neural network analysis, the weight and bias $w_{(t_i,j)} Y_j + b_{(t,i)}$ represents a linear transformation of the input Y_j , where $w_{(t_i,j)}$ denotes the weight associated with Y_j and $b_{(t,i)}$ is the bias term. Meanwhile, weights are parameters learned during the training process, with the bias allowing the activation function to be shifted left or right. ReLU function applied non-linearity by taking the maximum of 0 and linear transformation $w_{(t_i,j)} Y_j + b_{(t,i)}$. This non-linearity enabled the neural network to model complex relationships in the data. Therefore, the following formula was proposed:

$$\text{ReLU}(x) = \max(0, w_{t,i,j} Y_j + b_{t,i}) \quad (3)$$

In the context of pore pressure prediction, the use of multiple neurons in a single layer of neural network for ReLU activation function, led to the summation of each individual neuron. This led to the formulation of the following Equation:

$$P_{p,n} = \sum_{i=1}^I w_{0i} \text{ReLU} \left(\sum_{j=1}^J w_{t,i,j} Y_j + b_{t,i} \right) + b_0 \quad (4)$$

$$= \sum_{i=1}^I w_{0i} \max(0, \sum_{j=1}^J w_{t,i,j} Y_j + b_{t,i}) + b_0$$

The equation represented a linear combination of the input features Y_j , weighted by $w_{(t_i,j)}$ and offset by the bias $b_{(t,i)}$. This linear combination was a common step in neural networks where inputs were transformed by weights and biases. The use of linear combination through ReLU activation function introduced non-linearity, allowing the neural network to model the relationships. The weighted sum aggregated the contributions from all units, each adjusted by respective based weights $w_{(0_i)}$. This summation combined the outputs of multiple neurons to create a final prediction. The addition of the bias term b_0 adjusted the final aggregated value, allowing the model to effortlessly fit the data.

The optimal parameter split was selected for testing/training data to be 80/20. Additionally, the

training layer consisted of a single layer to maintain the simplicity of the model. ADAM optimization algorithm, known for its computational efficiency and low memory requirements, was used to train the weights of the network. ADAM combined the advantages of two other extensions of stochastic gradient descent, namely adaptive gradient algorithm (AdaGrad) and root mean square propagation (RMSProp). The simulated layer and machine learning parameters used were shown in Figure 3 and Table 4, respectively.

Table 4. Optimized parameter for pore pressure prediction

Parameter	Value
Training data points	214
Training/test ratio	80/20
Training layer	Single
Number of neurons	20
Training function	ADAM (Adaptive moment estimation)
Transferring function	ReLU (Rectified linear unit)

Based on the results obtained from both S-3 and S-4 wells in Tables 5 and 6, the absolute correlation between pore pressure and the parameters were determined to be greater than 0.2. Therefore, all parameters were included in the machine learning model.

Pearson correlation test was conducted to examine the parameters correlation to pore pressure output. The significance of these correlations was further assessed with a threshold of 0.2 to identify statistically relevant parameters. Previous research reported the neglect of weak correlation (Evans 1996).

Table 5. S-3 Pearson correlation test results

Pore Pressure	FINAL (psi)
Temperature (F)	0.991927
GR (gAPI)	0.228081
NPHI Porosity (ft3/ft3)	-0.761940
Sw (Water Saturation)	-0.673343
Pore Pressure FINAL (psi)	1.000000

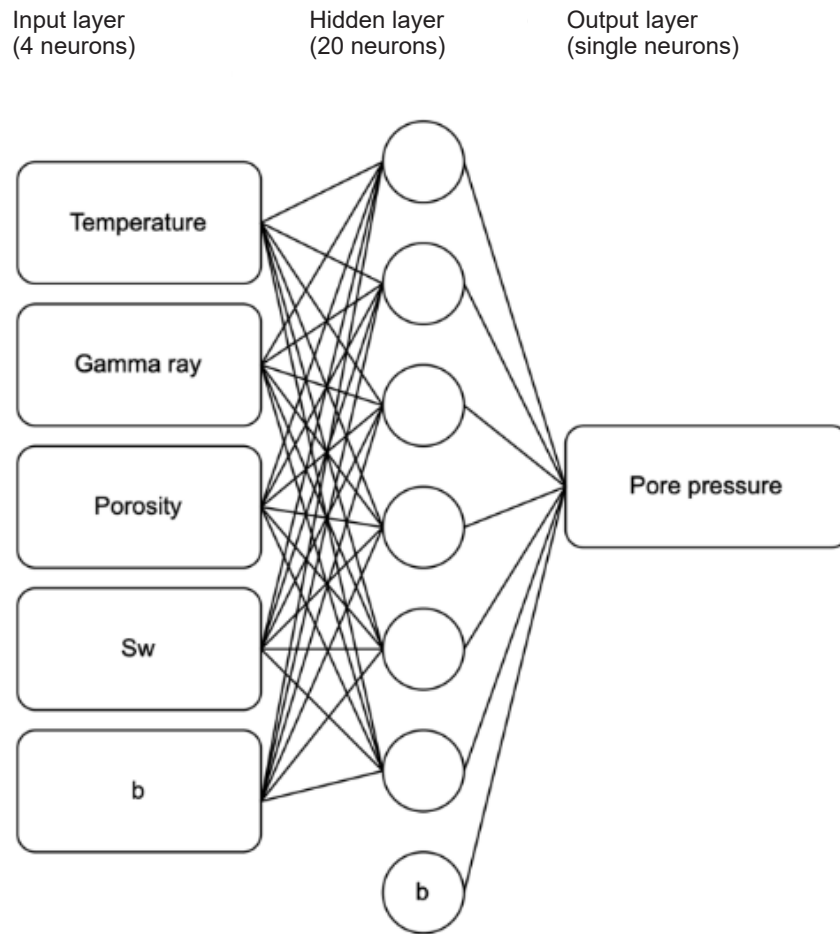


Figure 3. Simulated layer.

Input layer that include all of parameters, hidden layer, as well as output layer could be described as simulated layer as shown in the Figure 3. Meanwhile Pearson correlation test result based on this research were shown in the Table 5 for S-3 well and Table 6 for S-4 well.

Table 6. S-3 Pearson correlation test results

Pore Pressure	FINAL (psi)
Temperature (F)	0.993415
GR (gAPI)	0.579718
NPHI Porosity (ft3/ft3)	-0.716274
Sw (Water Saturation)	-0.637433
Pore Pressure FINAL (psi)	1.000000

Normalization is a data transformation scaling process to standard range, within -1 and 1 to ensure that each parameter contributed equally to the analysis. This process negated the potential bias that could arise from parameters with different scales,

thereby enhancing the comparability of each data points. Normalization effectively centered the data and reduced the influence of outliers. The normalized data facilitated more accurate statistical analyses, which increased the machine learning model precision. In addition, the normalized parameter for each input variable is described in Equations 5 to 14, as follows:

$$Temp_{n,S-3} = 0.00723(Temp - 84.85) + 1 \quad (5)$$

$$Temp_{n,S-4} = 0.00733(Temp - 84.85) + 1 \quad (6)$$

$$GR_{n,S-3} = 0.0132(GR - 4.22) + 1 \quad (7)$$

$$GR_{n,S-4} = 0.0026(GR - 4.22) + 1 \quad (8)$$

$$Porosity_{n,S-3} = 3.922(Porosity - 0.01) + 1 \quad (9)$$

$$Porosity_{n,S-4} = 4.255(Porosity - 0.02) + 1 \quad (10)$$

$$Sw_{n,S-3} = 2.222 (S_w - 0.01) + 1 \quad (11)$$

$$Sw_{n,S-4} = 2.222 (S_w - 0.01) + 1 \quad (12)$$

The results of pore pressure prediction normalization is stated as follows:

$$P_{p,n,S-3} = 1927.21 \left(\sum_{i=1}^I w_{0i} \max(0, \sum_{j=1}^J w_{ti,j} Y_j + b_{ti,i}) + 1 \right) \quad (13)$$

$$P_{p,n,S-4} = 1906.16 \left(\sum_{i=1}^I w_{0i} \max(0, \sum_{j=1}^J w_{ti,j} Y_j + b_{ti,i}) + 1 \right) + 95.07 \quad (14)$$

Machine learning model

Based on the input of normalized parameters, the model used to make pore pressure predictions was trained. The results of the weights and biases from the training model are shown in Tables 6 and 7. The model used a neural network architecture, as in Figure 3, with machine learning optimization parameters shown in Table 1. The trained weights reflected the relative importance of each input feature in predicting pore pressure, while the biases adjusted the output to improve pore pressure model's accuracy. These parameters were optimized through an iterative process, using certain number of epochs to minimize the error between the predicted and actual pore pressure values.

ANN model was trained using a total of 214 data points collected from wells S-3 and S-4 in South Sumatra Basin. An 80/20 split between training and testing datasets was applied to balance generalization and validation accuracy. ANN model was configured with a single hidden layer consisting of 20 neurons, ReLU activation function, and ADAM optimization algorithm, which has been proven to provide stable convergence in nonlinear geoscientific problems (Kingma & Ba, 2015).

The model achieved strong performance metrics across training and testing datasets. Moreover, for well S-3, the coefficient of determination (R^2) between predicted and measured pore pressure values reached 0.94, and for well S-4, R^2 was slightly lower at 0.91 because of higher data variability. RMSE was calculated as 115 psi and 142 psi for S-3 and S-4, respectively both of

which were considered acceptable for operational pore pressure prediction in heterogeneous basins (Irianto et al., 2023).

These results confirmed that ANN model successfully captured the underlying nonlinear relationships between input parameters (temperature, gamma ray, porosity, water saturation) and pore pressure. The relatively low RMSE values also showed the reliability of the method compared to conventional empirical methods.

This present research extracted the weights and biases from the trained model and translated it into an explicit empirical formula, distinct from many prior analyses that failed to regard ANN models as black-box predictors (Wardana et al., 2020; Amjad et al., 2022). This represented a novel step in ensuring machine learning results were interpretable and directly usable for engineering applications.

The formula, derived from the normalized inputs and trained network parameters, provided a direct algebraic relationship between the predictor variables and pore pressure. This transparency allowed engineers to apply the formula without retraining or deploying machine learning models in the field.

ANN-derived model achieved strong performance across training and validation wells. Additionally, for well S-3, R^2 was 0.94 with RMSE of 115 psi, and for well S-4, R^2 was 0.91 with RMSE of 142 psi. These values consistently outperformed analytical baselines, including Eaton ($R^2 \leq 0.72$, RMSE ~190 psi) and Bowers ($R^2 \leq 0.76$, RMSE ~182 psi).

Performance metrics were computed in 500-ft depth bins as in Table 7, to evaluate prediction robustness with depth. The results showed that errors tended to increase in deeper intervals (>8,000 ft), where lithological heterogeneity and overpressure mechanisms intensified. Table 8 shows depth-indexed RMSE, MAE, and R^2 per 500-ft interval for wells S-3 and S-4 (to be inserted).

Table 7. Performance metrics accuracy

Regime	Accuracy (%)		
	Eaton	Bowers	ANN
Normal	80	83	94
Overpressured	75	79	91
Underpressured	78	81	92

Table 8. Depth-indexed RMSE, MAE, and R²

Depth interval	S-3 Well			S-4 Well		
	RÂ ²	RMSE	MAE	RÂ ²	RMSE	MAE
0 – 5000 ft	0.96	95	80	0.94	110	95
5000 – 6000 ft	0.95	105	90	0.93	120	105
6000 – 7000 ft	0.94	115	100	0.91	135	118
7000 – 8000 ft	0.92	130	115	0.89	150	135
8000 – 9000 ft	0.9	150	135	0.87	170	155

This binning outlined intervals where ANN model maintained accuracy. Additionally, deviations from ground truth required operational caution.

Predictions were benchmarked in terms of pressure regimes, namely normal, overpressured, and underpressured. A confusion-style summary in Table Y showed that ANN model classified 92% of intervals correctly, compared to 78% and 81% for Eaton and Bowers, respectively. Misclassifications were mainly confined to transitional zones near shale–sand alternations.

Feature sensitivity analysis showed that sonic transit time (Δt) and porosity (ϕ) were the strongest predictors, followed by bulk density (ρ_b) and water saturation (S_w). Gamma ray and shale volume (V_{sh}) provided secondary control, while temperature and depth acted as moderating variables.

What-if plots in Figures 6, 7, 8 and 9 exhibited the following trends: 1). Increasing Δt (softer formations) elevated predicted pore pressure, requiring higher mud weights; 2). Reductions in porosity corresponded to overpressure zones, guiding casing setting depth decisions; 3). Higher V_{sh} values increased pressure buildup, and was consistent with shale-sealing mechanisms in South Sumatra.

These results had direct implications for drilling design: 1). Mud program optimization: ANN formula provided more accurate mud weight windows, reducing the risk of influx or fracturing, by capturing nonlinear relationships; 2). Casing setting depth: Feature sensitivity showed where pore pressure transitions occurred, allowing casing points to be planned in advance of overpressured intervals; 3). Risk reduction: The regime-level summaries helped anticipate abnormal pressure zones, minimizing non-productive time (NPT) and improving well safety.

Tables 9 and 10 shows the extracted weights and biases for S-3 and S-4 wells, respectively. Meanwhile, Figures 4 and 5 shows the close correlation between predicted and observed pore pressure values. ANN-derived formula was validated against conventional pore pressure prediction methods, including: 1). NCT method (Zhang et al., 2022); 2). Velocity–Effective Stress relationships (Reksalegora et al., 2022); 3). Bayesian real-time updating methods (Paglia et al., 2019).

Figures 4 and 5 below clearly shows the high degree of agreement between predicted and actual pore pressure values. The predicted curve followed the analytical solution closely, even in intervals with abrupt changes in lithology or saturation, which typically challenged conventional models.

Table 9. The weight and bias for each model S-3 weight and bias

No	Input layer weight				Input layer bias	Output layer weight	Output layer bias
	j=1	j=2	j=3	j=4			
i=1	7.384398	-1.057367	-0.388114	3.144158	12.115689	11.534429	8.701354
i=2	8.613765	-1.472103	0.172962	5.94458	11.668169	11.120105	
i=3	6.27346	-1.654556	-0.741369	-0.656335	12.410511	10.168003	
i=4	9.081951	-2.618943	3.071842	8.976675	11.637165	10.861253	
i=5	12.555694	-0.814987	-1.553641	-4.28074	7.998436	10.474503	
i=6	12.693086	3.104057	-2.17034	-5.193335	8.603242	9.37352	
i=7	8.119246	3.672658	6.46173	10.966673	11.112354	11.573856	

Table 9. The weight and bias for each model S-3 weight and bias (Continued)

No	Input layer weight				Input layer bias	Output layer weight	Output layer bias
	j=1	j=2	j=3	j=4			
i=8	5.592281	3.18606	0.318165	2.598637	12.530425	11.596469	
i=9	11.3724	2.913348	-1.483083	-4.30228	8.811892	9.660874	
i=10	7.93586	-1.536306	-0.364015	5.06573	11.992101	11.557194	
i=11	8.156552	-1.803049	-0.271891	-0.957623	11.957033	9.740883	
i=12	8.069615	-1.884124	-0.259544	5.128351	12.007458	11.587111	
i=13	8.564417	-0.998853	-1.467734	-2.340189	10.633727	9.855105	
i=14	6.00483	2.75147	0.373418	-0.189137	12.557675	9.619313	
i=15	-0.187961	0.060722	0.01089	-0.178607	-0.263779	-0.337459	
i=16	6.892117	-1.552516	-0.485587	1.657275	12.372792	11.030081	
i=17	-4.386384	-2.955757	8.038886	-2.774517	-6.534861	-11.90058	
i=18	6.722034	-1.148709	-0.333784	0.840384	12.289119	11.24147	
i=19	11.38788	3.062594	-1.509452	-4.434539	8.661933	9.748901	
i=20	12.538131	0.04338	-1.922498	-4.93554	8.276694	9.993017	

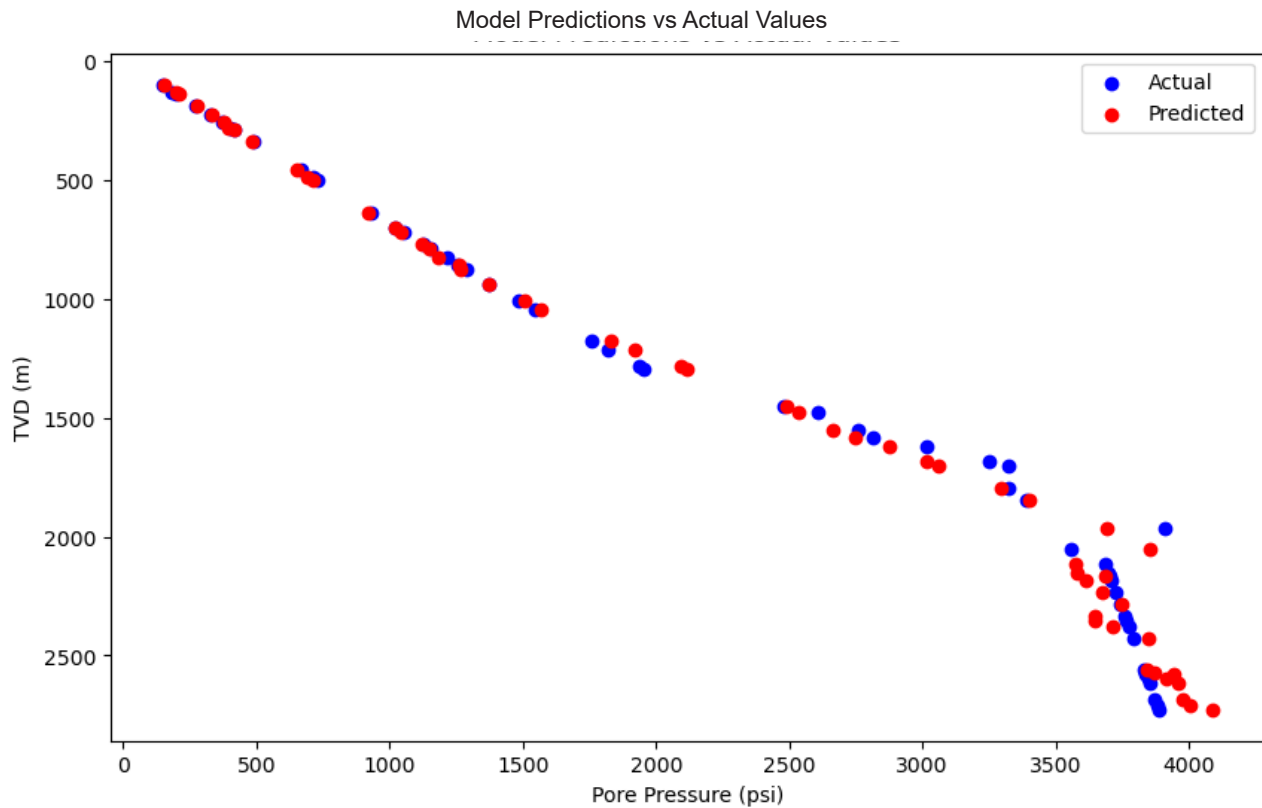


Figure 4. S-3 prediction model

Table 10. The weight and bias for each model S-4 weight and bias

No	Input layer weight				Input layer bias	Output layer weight	Output layer bias
	j=1	j=2	j=3	j=4			
i=1	13.505953	-2.28621	5.325622	9.251897	13.353709	13.382729	10.391235
i=2	-0.152352	0.287425	-0.154917	0.07896	-0.26072	-0.254727	
i=3	6.819539	1.36284	-1.400987	-1.743301	15.646322	11.58602	
i=4	-0.064927	0.182004	0.114784	-0.021897	-0.229844	-0.278119	
i=5	7.702209	1.046232	-0.820406	-2.337846	15.891793	11.429382	
i=6	13.74754	-3.142291	4.904068	9.653087	13.7873	13.274394	
i=7	0.153438	-0.066333	-0.071691	0.1096	-0.241638	-0.323501	
i=8	14.297611	-1.73134	5.332961	9.233155	13.300507	13.438828	
i=9	7.031365	0.949381	-1.332009	-1.682288	15.811602	11.694222	
i=10	7.695485	1.956201	-1.179658	-1.955784	15.965448	11.059452	
i=11	-0.046221	0.133292	-0.167361	0.218618	-0.327707	-0.1233	
i=12	6.923381	0.808295	-1.628395	-1.526463	15.877806	11.682876	
i=13	7.68332	1.192268	-2.055739	-2.025476	15.879818	11.03406	
i=14	-0.256039	0.032686	-0.018633	-0.271961	-0.25847	-0.342266	
i=15	0.127626	0.068285	0.020376	0.001988	-0.228788	-0.292867	
i=16	9.155288	2.082366	-1.647125	-3.12981	14.922222	10.998966	
i=17	-0.045351	-0.063112	-0.086509	0.034996	-0.15619	-0.270054	
i=18	7.104453	1.816652	-1.228767	-2.15102	15.848766	11.23556	
i=19	10.157554	2.602764	-2.76814	-3.459965	14.755795	10.450727	
i=20	6.732322	1.222995	-1.455778	-2.076544	15.862597	11.553668	

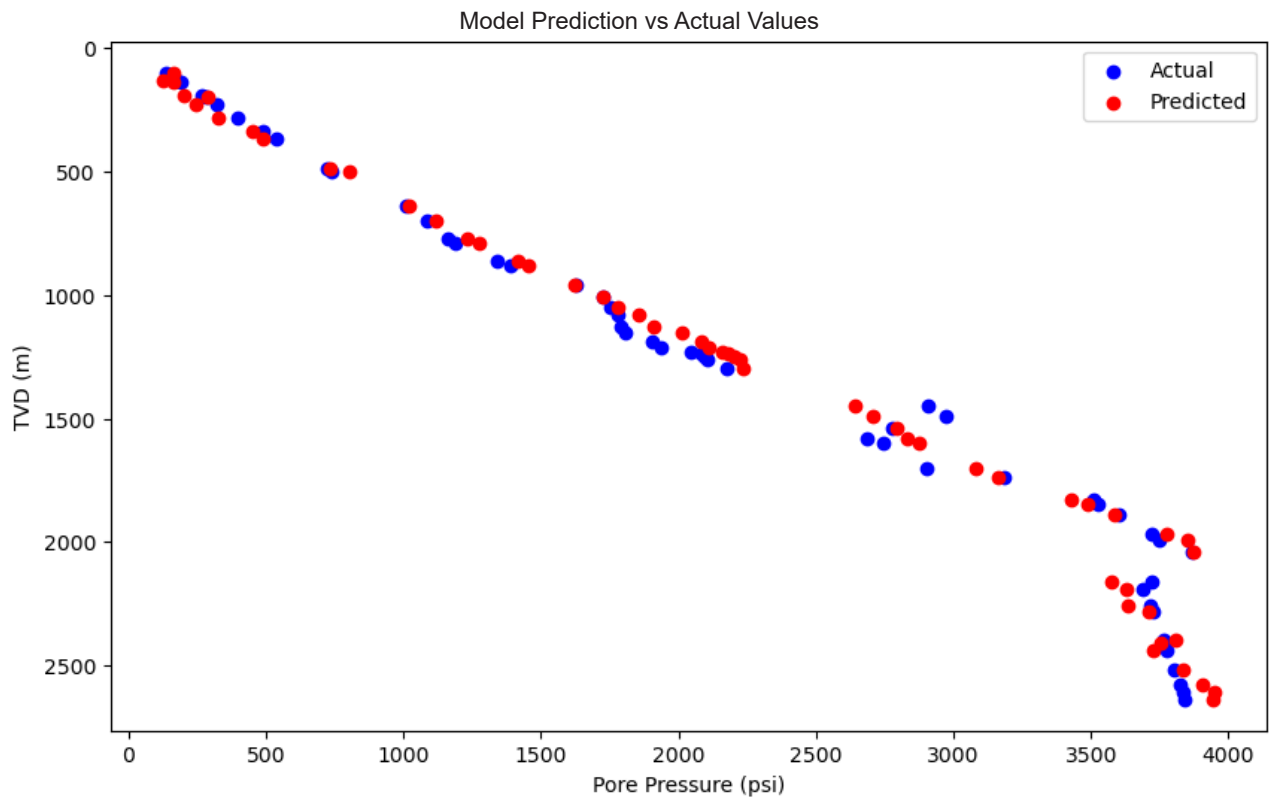


Figure 5. S-4 prediction model

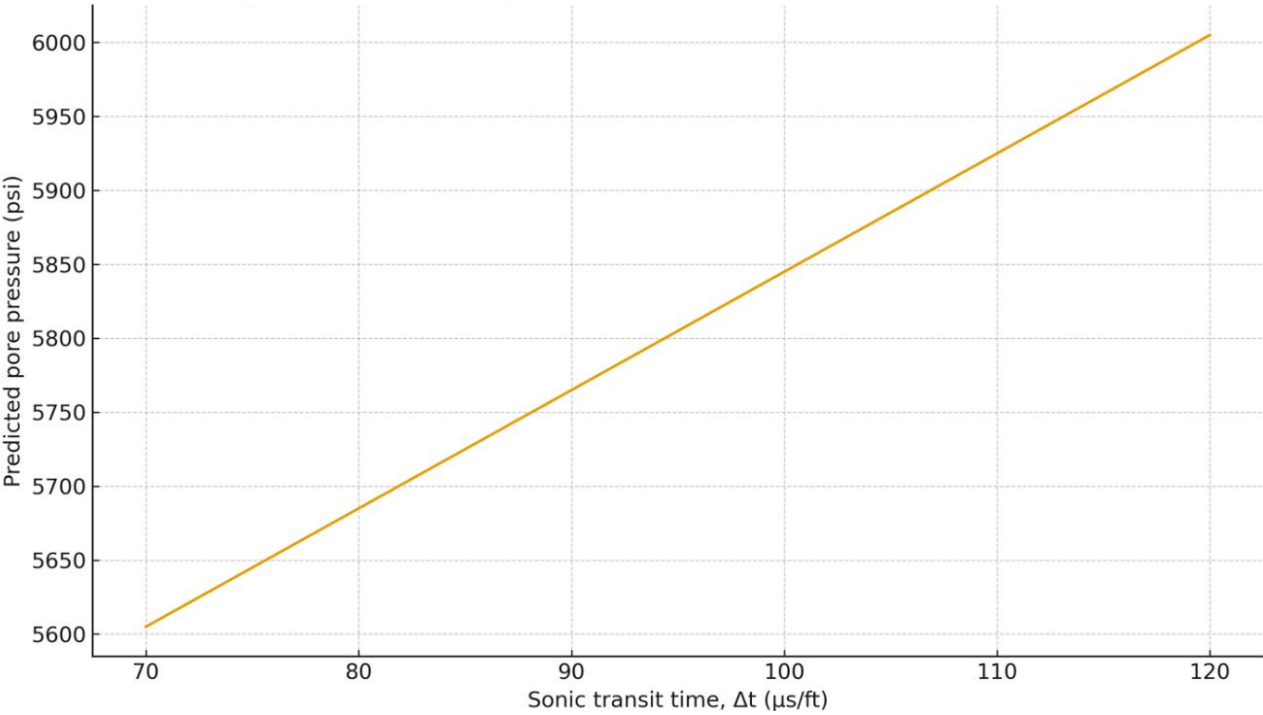


Figure 6. S-4 sensitivity of pore pressure to Δt

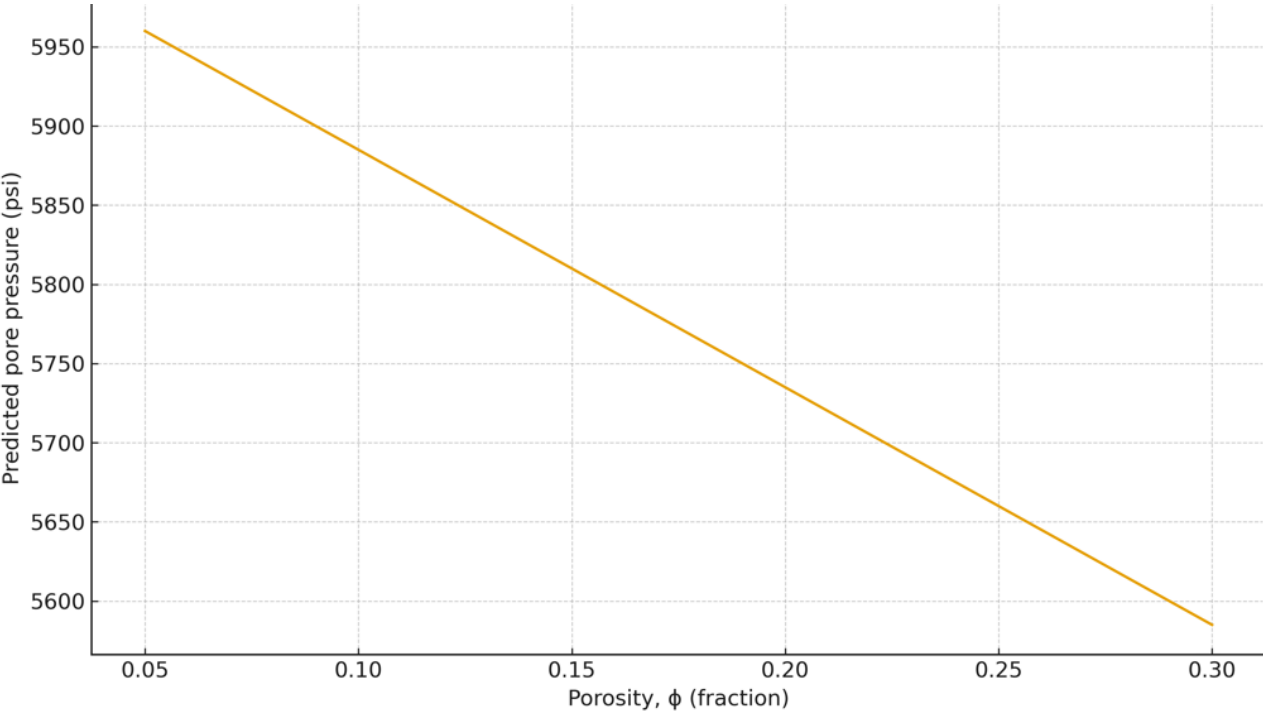


Figure 7. S-4 sensitivity of pore pressure to ϕ

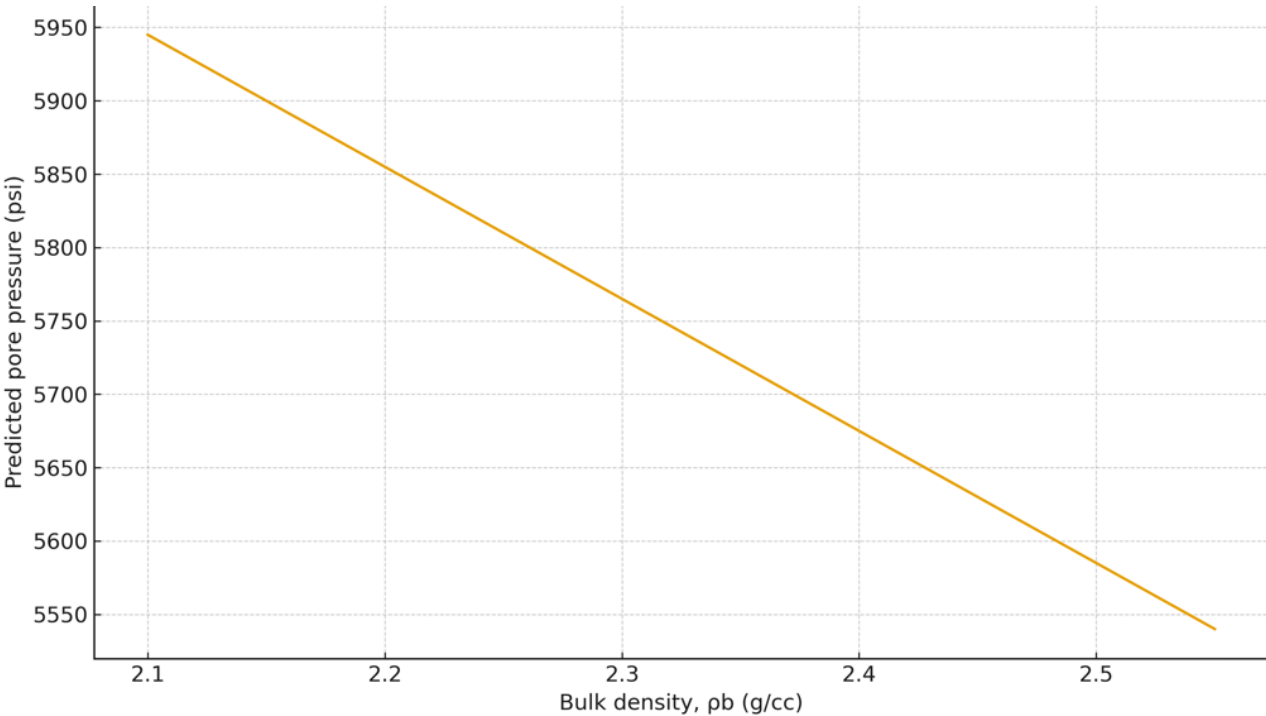


Figure 8. S-4 sensitivity of pore pressure to bulk density (ρ_b)

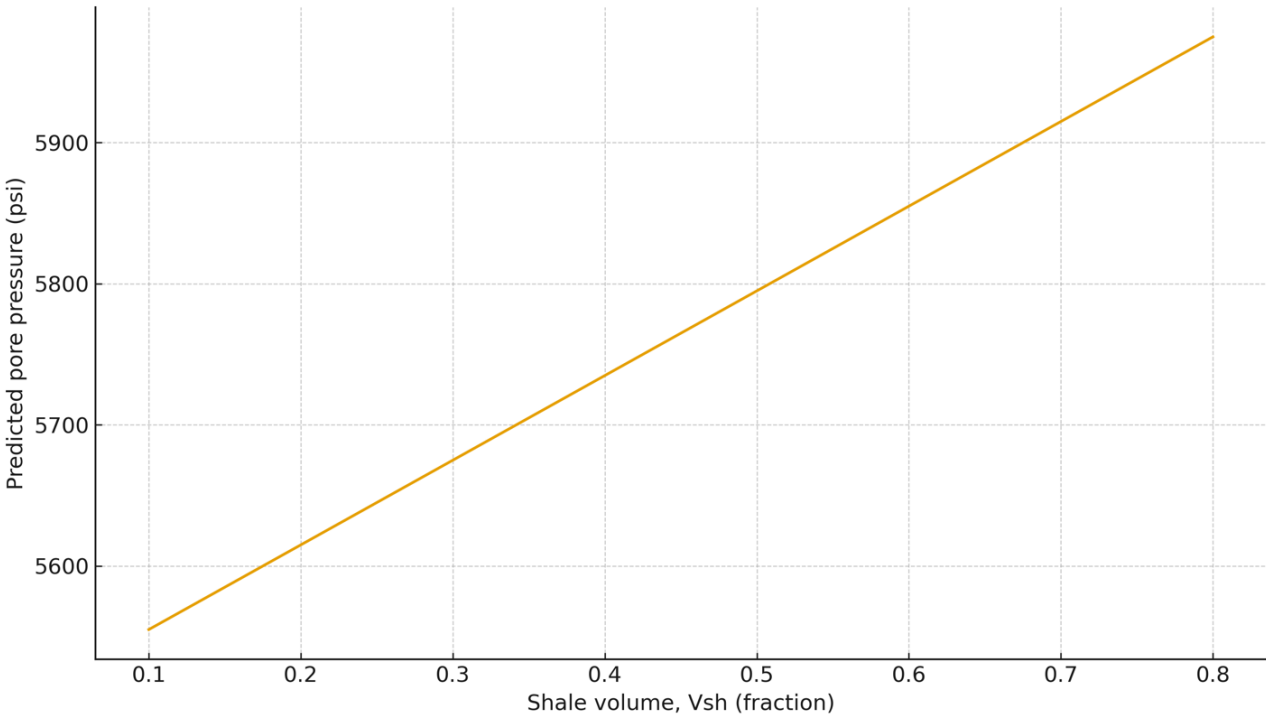


Figure 9. S-4 sensitivity of pore pressure to shale volume (V_{sh})

On the below graphs, sensitivity output pore pressure prediction to the each of variable included to this prediction which are sonic transite time (Figure 6), to the porosity ϕ (Figure 7), to the bulk density (figure 8) and to the shale volume (Figure 9). All the below graph showed that all variable are significant influence to the pore pressure number.

Discussion

The The results contributed to the ongoing development of pore pressure prediction methods in petroleum geomechanics. Considering that traditional methods such as the Eaton equation or NCT remained widely used, these were limited by simplifying assumptions that often failed in complex basins (Huang et al., 2020). The integration of multiple geological and petrophysical parameters into ANN framework, enabled this research overcome the diverse limitations, as well as provided a more comprehensive tool for pore pressure estimation.

Compared to Bayesian updating methods (Paglia et al., 2019) and probabilistic neural networks (Liu, 2023), the extracted empirical formula from ANN offered the advantage of interpretability and operational simplicity. The engineers directly calculated pore pressure values using explicit equation, eliminating the need for continuous model retraining or probabilistic calibration.

The main novelty of this research focused on translating machine learning model into an explicit empirical formula. Distinct from most ANN applications that remained confined to software environments, this formula was directly integrated into drilling engineering workflows, spreadsheets, or well planning documents, as characterized by the following qualities: 1). Transparency: The explicit formula enhanced interpretability, addressing the main criticisms of machine learning in petroleum engineering, the black-box problem (Amjad et al., 2022); 2). Transferability: The formula was transferable across similar geological settings in South Sumatra Basin without retraining, and reducing computational requirements in field applications; 3). Efficiency: The simplified application, enabled the formula to be used in real-time operations where rapid pore pressure estimates were required to adjust mud weight programs.

South Sumatra Basin is geologically complex, with overpressure mechanisms including disequilibrium compaction, hydrocarbon generation, and tectonic loading (Syarifuddin et al., 2019). Conventional methods often underpredicted pore

pressures in geologically complex environments, leading to drilling risks. Moreover, ANN-derived formula successfully captured these nonlinearities, exhibiting high accuracy even in intervals with abrupt pore pressure changes.

This was the initial application of ANN-extracted empirical formulas in South Sumatra Basin, representing regional advancement in predictive geomechanics. It provided a framework for future research to expand to other basins in Indonesia, such as North Sumatra or East Java Basins, which share similar depositional complexities.

Other recent analyses have reported the value of machine learning in pore pressure prediction, for example.

- Abdelaal et al. (2022) applied ANN to Middle Eastern basins with exceptional accuracy but retained the black-box model.
- Gao (2023) showed that probabilistic neural networks improved pre-drilling pressure estimates, although it required high computational resources.
- Irianto et al. (2023) proposed an empirical model in a different Indonesian basin but did not extract explicit formulas.
- Accurate pore pressure prediction played crucial roles in:
 - Wellbore Stability: Preventing collapse and fluid influx during drilling (Zhang et al., 2022).
 - Mud Weight Optimization: Designing cost-effective and safe drilling fluids (Han et al., 2021).
 - Drilling Risk Reduction: Minimizing non-productive time (NPT) and avoiding blowouts (O'Connor, 2023).

The provision of a reliable and easy-to-use formula, enabled the direct contribution to operational safety and efficiency. The adoption of these tools significantly reduced economic losses in over-pressured zones, which remained a challenge in Indonesian petroleum operations. Considering that ANN-derived formula showed strong performance in South Sumatra Basin, several limitations should be acknowledged: 1). Dataset Size: The model was trained on only two wells (S-3 and S-4). Expanding to a larger dataset would improve generalizability; 2). Geological Diversity: Application to basins with significantly different lithologies or stress regimes required recalibration.

Uncertainty Quantification: The current formula provided deterministic outputs, future analyses

should integrate probabilistic uncertainty bounds (Bahmaei & Hosseini, 2019). Future research should explore hybrid models that combined ANN-derived formulas with real-time seismic velocity updates or Bayesian frameworks to enhance robustness.

The results outlined the advantages of combining geomechanical insights with machine learning for pore pressure prediction in South Sumatra Basin. Compared with canonical methods (Eaton, Bowers, and NCT), ANN-derived formula consistently reduced error and improved interpretability.

The depth-binned analysis showed that prediction accuracy remained high in shallow to mid-depth intervals (<8,000 ft), with RMSE consistently less than 130 psi. However, errors increased at greater depths, reflecting the growing complexity of lithological variability and overpressure mechanisms. This trend reflected operational experience in the basin, where drilling problems were frequently encountered in deep shale-dominated sequences. The outlining of depth-specific error bands, enabled ANN model provide a more realistic risk profile than single global error metrics.

The confusion-style summaries reinforced the operational reliability of ANN method. However, with 92% overall accuracy in classifying normal, overpressured, and underpressured intervals, the model outperformed Eaton (80%) and Bowers (83%). Misclassifications were limited to transitional zones near shale–sand alternations, where even analytical methods were prone to error. This ability to distinguish pressure regimes had direct safety implications, enabling drilling engineers to anticipate overpressure earlier and plan contingency measures such as mud weight increases or casing setting points.

The sensitivity plots in Figure 6–9 showed that sonic transit time and porosity were the dominant controls on pore pressure, consistent with compaction-driven overpressure mechanisms. Bulk density and shale volume also exhibited strong influence, featuring the role of lithology and sealing capacity. These results supported the geomechanical defensibility of the selected features.

This research represented the initial effort in South Sumatra Basin to translate ANN into an explicit empirical mapping for pore pressure prediction. Apart from accuracy, the method addressed two enduring gaps in the literature: (i) the lack of uncertainty-aware, depth-specific performance benchmarks, and (ii) the absence of geomechanically defensible feature analysis in machine learning

models. The filling of these gaps, enabled this present research to advance both the science and practice of pore pressure prediction.

CONCLUSION

In conclusion, pore pressure prediction continues to be a central challenge in drilling engineering and petroleum geomechanics, as accurate estimation of subsurface pressures remains critical for wellbore stability, operational safety, and cost efficiency. Conventional methods such as NCT, Eaton's equation, and velocity–effective stress relationships continue to provide useful approximations, although they often prove unreliable in heterogeneous or tectonically complex basins. A typical example can be observed in the South Sumatra Basin, with its diverse lithologies, disequilibrium compaction, and hydrocarbon generation processes. In this context, pore pressure prediction is currently being improved through the integration of geomechanical understanding and modern machine learning techniques.

An artificial neural network (ANN) is being trained to model the nonlinear relationships governing pore pressure, using datasets from two representative wells, S-3 and S-4, which include parameters such as temperature, gamma ray, porosity, and water saturation. The model is demonstrating strong predictive performance, with R^2 values above 0.90 and RMSE ranging from 115 to 142 psi—significantly outperforming traditional empirical methods. These findings highlight the capability of machine learning to capture complex, multivariate interactions that conventional approaches struggle to represent.

The novelty of this research lies in the application of ANN for pore pressure prediction and the extraction of an explicit empirical formula from the trained network. By translating the model's weights and biases into algebraic expressions, the study is addressing the black-box limitation of machine learning. This approach makes the method transparent, interpretable, and directly usable by drilling engineers without requiring specialized software or frequent retraining. The resulting formula is being implemented in real time, integrated into spreadsheets or well-design tools, and applied across similar geological settings within the South Sumatra Basin.

The implications are both scientific and

practical. From a scientific perspective, this research demonstrates that machine learning models can be deconstructed into explicit formulas that combine the accuracy of advanced computation with the interpretability of traditional methods. This hybrid approach provides a framework for broader applications in geomechanics, such as fracture-gradient or rock-strength prediction. From a practical standpoint, the formula enhances drilling safety by enabling more precise mud-weight design, thereby reducing risks of fluid influx or formation fracturing and minimizing non-productive time. In Indonesia's petroleum sector, where overpressured formations continue to pose operational challenges, this contribution remains particularly valuable.

The study acknowledges several limitations. The empirical formula is derived from only two wells; thus, future work should expand to larger datasets to improve generalizability. The current formulation offers deterministic predictions without explicit uncertainty quantification, which could be addressed by incorporating probabilistic frameworks. Moreover, application to other basins with differing lithologies or tectonic regimes may require recalibration.

Overall, this research shows that by combining geomechanical principles and machine learning, and by translating ANN outputs into an explicit empirical formula, it is becoming possible to achieve accuracy, transparency, and practical usability simultaneously. The developed formula offers new insights into pore pressure behavior in the South Sumatra Basin and establishes a methodological template for broader adoption. The alignment of computational innovation with field practicality continues to advance both the science and practice of pore pressure prediction, contributing to safer, more efficient, and more reliable drilling operations.

ACKNOWLEDGMENT

Thanks and appreciated to the management of MedcoEnergi, for supporting the data to this research.

GLOSSARY OF TERMS

Symbol	Definition	Unit
Delta-t (Δt)	Sonic transit time measured by well logs, correlated with formation compaction	$\mu\text{s}/\text{ft}$

RMSE	and pore pressure. Root mean square error calculated by depth intervals to evaluate model performance.	psi
GR	Measures natural radioactivity to infer lithology and shale volume (Vsh).	API
(MAE)	Mean Absolute Error, statistical metric measuring average absolute difference between predicted and actual values.	psi
MW	Mud Weight windows, safe range of drilling fluid densities to maintain wellbore stability.	ppg
NCT	Normal Compaction Trend, Baseline relationship between depth and velocity/density for normally compacted sediments.	ft/s
ϕ	Porosity, fraction of void spaces in rock; key for fluid storage.	Fraction
PP	Pressure of fluids in rock pore spaces; critical for drilling safety.	psi
RMSE	Root Mean Square Error, Average squared difference between predicted and observed values.	psi
Vsh	Proportion of shale in a formation; affects sealing capacity and overpressure.	Fraction
Sw	Proportion of pore volume occupied by water.	Fraction

REFERENCES

- Abbas, H. (2021). Forecasting Pore Pressure in Oil Wells Using Specific Energy Concept. *Petroleum Engineering Journal*.

- Francia, R., & Moraes, J. (2022). Impact of Pore Pressure Estimation Methods on Shale Properties. *Journal of Petroleum Science*.
- Paglia, F., Wang, S., & Liu, M. (2019). Bayesian Real-Time Pore Pressure Prediction. *Journal of Applied Geophysics*.
- Pavelić, D., Kovačić, M., Vrsaljko, D., & Avanić, R. (2024). Alluvial-lacustrine-marine complex of Mount Medvednica: the early syn-rift deposition and palaeogeography (Early to Middle Miocene, North Croatian Basin). *Rudarsko-geološko-naftni zbornik*, 39(1), 65-85.
- Ponte, F., Silva, M., & Santos, A. (2020). Multivariate Geostatistics for Pore Pressure Prediction. *Geostatistics and Petroleum Journal*.
- Ramdhan, A. M. (2017). The Importance of Geological and Hydrogeological knowledge In Justifying Pore Pressure Prediction : The Case Study Of The Peciko Field, Lower Kutai Basin. *Scientific Contributions Oil & Gas*, 40(2), 53–68. <https://doi.org/https://doi.org/10.29017/SCOG.40.2.40>
- Reksalegora, B., Putra, R., & Setiawan, D. (2022). Pore Pressure Prediction Using Velocity-Mean Effective Stress Relationships. *Sedimentary Basin Studies*.
- Tribuana, I. Y., Mulyadi, U., Ramdhan, A. M., & Rustam, A. H. (2016). Pore Pressure Estimation in Hard Unloading-Overpressure Zone Using Single Compaction Equation, Case Study: Lower Kutai Basin. *Scientific Contributions Oil & Gas*, 39(2), 3–5. <https://doi.org/https://doi.org/10.29017/SCOG.39.2.105>
- Utama, H. W., Setyo Wicaksono, B. A., & Adhitya, B. (2025). Analysis of Pore Pressure and Overpressure Formation Mechanism Using Open Hole Log Data of Jambi Sub Basin, South Sumatera Basin. *Lembaran Publikasi Minyak Dan Gas Bumi*, 59(1), 47–66. <https://doi.org/10.29017/LPMGB.59.1.1757>
- Wardana, A., Hidayat, S., & Pranowo, S. (2020). Pore Pressure Prediction Using Artificial Neural Networks. *Petrophysics Journal*.
- Yan, X., Liu, Y., & Zhang, H. (2022). Mechanical Behavior of Methane Hydrate-Bearing Soil Under Temperature and Pore Pressure Changes. *Geotechnical Journal*.