Scientific Contributions Oil & Gas, Vol. 49. No. 1, March: 351 - 366

SCIENTIFIC CONTRIBUTIONS OIL AND GAS

Testing Center for Oil and Gas
LEMIGAS

Journal Homepage:http://journal.lemigas.esdm.go.id
ISSN: 2089-3361, e-ISSN: 2541-0520

Prediction of S-Wave Using Conventional Method and Machine Learning

Dayyan Dhaifullah[1], Sonny Winardhi[1,2], and Ekkal Dinanto[2]

[1]Geophysical Engineering Graduate Program, Faculty of Mining and Petroleum Engineering , Institut Teknologi Bandung
Ganesa Street No. 10, Bandung, West Java, 40132, Indonesia.

[2]Seismological Engineering and Exploration Research Group, Faculty of Mining and Petroleum Engineering,
Institut Teknologi Bandung
Ganesa Street No. 10, Bandung, West Java, 40132, Indonesia.

Corresponding author: Dayyan Dhaifullah (22323004@student.itb.ac.id)

Manuscript received: January 30th, 2026; Revised: February 26th, 2026
Approved: February 27th, 2026; Available online: March 20th, 2026; Published: March 20th, 2026.

ABSTRACT - Shear-wave velocity (Vs) is an essential metric for subsurface characterization and CO_2 storage evaluation. However, Vs measurements are frequently unavailable in mature fields due to limited data acquisition. This research employs a machine-learning approach utilizing a Fully Connected Neural Network (FCNN) to predict Vs and Vp/Vs logs at a potential CO_2 injection site within a heterogeneous carbonate reservoir. Seismic elastic properties, particularly Acoustic Impedance (AI) and Vp/Vs, play a crucial role in assessing reservoir capacity by linking elastic responses to petrophysical properties such as porosity and water saturation. Conventional approaches, including the Castagna empirical relationship and Multiple Linear Regression (MLR), are commonly used for Vs estimation. Nevertheless, these methods often inadequately account for fluid-related effects. To address this limitation, this study examines two predictive approaches: (1) indirect Vp/Vs derived from predicted Vs, and (2) direct prediction of Vp/Vs prediction using a FCNN model. The findings indicate that direct Vp/Vs prediction demonstrate stronger correlation with observed data (R = 0.8023) and improved sensitivity to lithological and fluid variations compared to traditional methods. These findings underscore the advantage of directly predicting fluid-sensitive elastic properties through machine learning, providing a more reliable framework for reservoir characterization and CO_2 storage assessment in data-constrained carbonate formations.

Keywords: machine learning, Vp/Vs, multi-linear regression, castagna.

How to cite this article:

Dayyan Dhaifullah, Sonny Winardhi, and Ekkal Dinanto, 2026, Prediction of S-Wave Using Conventional Method and Machine Learning, Scientific Contributions Oil and Gas, 49 (1) pp. 351-366. DOI org/10.29017/scog.v49i1.1988.

INTRODUCTION

This study aims to predict shear-wave velocity (Vs) at a site designated for potential CO_2 injection. Achieving global net-zero emission targets requires effective CO_2 mitigation technologies to reduce atmospheric greenhouse gas concentrations. In this context, seismic data are essential for monitoring CO_2 injection and modeling carbon storage using elastic attributes such as Acoustic Impedance (AI) and the Vp/Vs ratio. These elastic properties can be correlated with petrophysical parameters, including porosity and water saturation, to evaluate the reservoir's long-term carbon storage potential.

However, as commonly observed in mature fields designated for CO_2 injection, data availability remains limited particularly for shear-wave velocity (Vs). Vs is a critical parameter in CO_2 monitoring as it provides a more robust representation of subsurface rock properties and is relatively insensitivity to fluid variations, thereby reflecting intrinsic lithological characteristics such as porosity. Consequently, developing reliable techniques to compensate for the lack of Vs data is essential for improving reservoir characterisation, especially in heterogeneous carbonate reservoirs.

In recent years, the utilization of machine-learning methodologies in geophysical and petrophysical studies has grown substantially. Numerous studies have demonstrated that data-driven models outperform traditional empirical and linear approaches in predicting elastic properties by effectively capturing nonlinear relationships among well-log parameters. By integrating multiple petrophysical logs, machine-learning models achieve higher predictive accuracy and better generalization across complex lithologies, thereby improving reservoir characterization (Mehrad et al., 2022; Zhang et al., 2022). Machine Learning (ML) also provides a cost-effective and efficient alternative to conventional methods for log prediction (Zainuri, 2023). However, many previous studies have primarily focused on prediction accuracy without thoroughly addressing fluid sensitivity and its implication for interpreted elastic properties. Conventional techniques for estimating Vs, such as the Castagna empirical relationship and Multiple Linear Regression (MLR), have been widely employed and generally produce acceptable results. However, these approaches may struggle to accurately capture fluid -related effects, given that Vs is inherently less sensitive to fluid content variations. As a results, Vs predictions derived from fluid-sensitive input parameters may lead to misinterpretations in reservoir characterization and CO_2 monitoring.

To overcome these limitations, this study employs a machine-learning framework based on a Fully Connected Neural Network (FCNN) to predict Vs and directly estimate the Vp/Vs ratio. The FCNN, also known as a dense neural network, is one of the most widely used artificial neural network architectures, characterized by full connectivity between successive layers and its ability to model complex nonlinear relationships. Previous studies have successfully applied neural network methodologies for Vs prediction (Rajabi et al., 2023; Mousavi et al., 2024; Bettir et al., 2024), demonstrating superior performance compared to conventional approaches in heterogeneous reservoirs.

The novelty of this research lies in the comparative evaluation of two predictive strategies: (1) indirect estimation of Vp/Vs based on predicted Vs, and (2) direct prediction of the Vp/Vs ratio using an FCNN model. By comparing this strategies with traditional approaches such as MLR and the Castagna equation, this study emphasizes the benefits of directly predicting elastic properties that are more sensitive to fluid variations. This approach reduces uncertainty associated with fluid effects and provides a more reliable representation of subsurface conditions for CO_2 injection evaluation.

Overall, this framework contributes methodologically by demonstrating that direct machine-learning prediction of Vp/Vs exhibits greater sensitivity to lithological and fluid variations than traditional Vs-based methods. The findings provide significant insights for predicting elastic properties in data-scarce carbonate reservoirs and support the application of machine learning for improved reservoir characterisation in CO_2 storage research. By leveraging nonlinear relationships among well-log parameters, machine-

learning models can generate predictions that more closely represent actual subsurface conditions.

METHODOLOGY

In this study, Castagna method and machine learning is used for prediction of Vs by utilizing Vp/Vs approach. The Castagna method is use as bridge to help machine learning building a initial model for prediction of Vp/Vs, as the data in this study has a limitation on depth for complete well, Castagna method is used to help the deep well and learning the Vs log model for the machine learning initial model build. When the result shows a good correlation to the complete data well. The model will be used to predict the Vs data for well that have a missing Vs log.

Empirical method of castagna

This study applies the Castagna wet limestone relationship as a validation tool for wells without measured Vs data. Cross-well validation is essential in machine learning processes to ensure that the model's predictive performance accurately represents actual subsurface conditions. The Castagna wet limestone relationship is an empirical correlation developed by Castagna et al. (1993) to estimate shear-wave velocity (Vs) from compressional-wave velocity (Vp) in water-saturated carbonate rocks (wet limestone). This correlation was established using laboratory measurements and well-log data across various lithologies, including sandstone, shale, and water-saturated limestone. As illustrated in figure 1, a linear relationship exists between Vp and Vs. Using Equation (1), Vs logs can be derived from measured Vp logs. For wet limestone lithology, the empirical equation is expressed as follows:

$$Vs = -0.05508\,.\,V_P^2 + 1.0168V_P - 1.0305 \qquad (1)$$

This equation demonstrates that lithology and fluid saturation influence the relationship between shear-wave velocity (Vs) and compressional-wave velocity (Vp). Therefore, the Castagna wet limestone correlation serves as a reliable empirical reference for validating the Vs predictions generated by machine-learning models.

In addition, this study employs Multiple Linear Regression (MLR) to predict the Vs log, providing a benchmark for comparison with machine-learning results. Rather than only using Vp log as a conventional predictor of Vs, this study modifies the traditional approach by incorporating multi-attribute inputs within the MLR framework. The resulting Vs predictions derived from MLR are subsequently used to estimate either Vs directly or the Vp/Vs ratio.

Multi-linear regression.

Multi-Linear Regression (MLR) is a statistical technique employed to represent the linear relationship between a dependent variable (Y) and

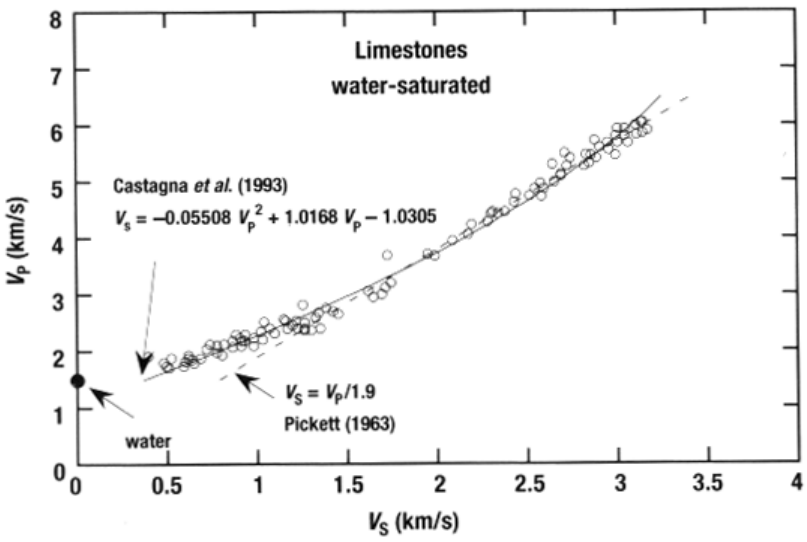

Figure 1: Crossplot between Vs and Vp illustrating the castagna wet limestone empirical relationship (Castagna et al., 1993).

multiple independent variables (X_1, X_2, ..., X_n). The MLR model aims to construct a linear equation that predicts the value of Y as a linear combination of the input variables. In geophysical log analysis, MLR is commonly applied in multi-attribute log prediction, where several well-log attributes are integrated to estimate missing or unmeasured log parameters. In this study, MLR is used as a baseline model to evaluate the predictive performance of Vs and Vp/Vs logs generated by the machine learning approach.

Fully connected neural network

The Fully Connected Neural Network (FCNN), also known as a dense neural network, is a fundamental and widely utilized design in Artificial Neural Networks (ANN). In a FCNN, every neuron in a given layer is directly connected to all neurons in the subsequent layer, unlike the localized connections employed in Convolutional Neural Networks (CNNs).

Figure 2 illustrates the structure of a typical FCNN neuron. This architecture emulates the way human brain processes information processing through interconnected layers of neurons. Each neuron computes its output as a weighted linear combination of its inputs, followed by the application of a nonlinear activation function.

The fundamental operation of a single fully connected layer can be mathematically expressed as follows:

$$z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)} \quad (2)$$

$$a^{(l)} = f\left(z^{(l)}\right) \quad (3)$$

where a(l - 1) is the output (activation) from the previous layer, W(^{l}) is the weight matrix that connects layer l - 1 to layer l, and b(^{l}) is the bias vector associated with layer l. In this study, FCNN method was utilized to predict log Vs, and also Vρ/ Vs ratio. The predictive performance of these two outputs is then compared to evaluate the effectiveness of direct versus indirect elastic property estimation.

A Fully Connected Neural Network (FCNN) architecture is developed based on the previously mentioned framework. The design of the model requires careful selection of several components known as hyperparameters, which are defined prior to the training process.

Unlike model parameters, which are learned directly from the data, hyperparameters are predefined and play a critical role in controlling model complexity, convergence behaviour, and generalization capability (Sudrazat et al. 2020).

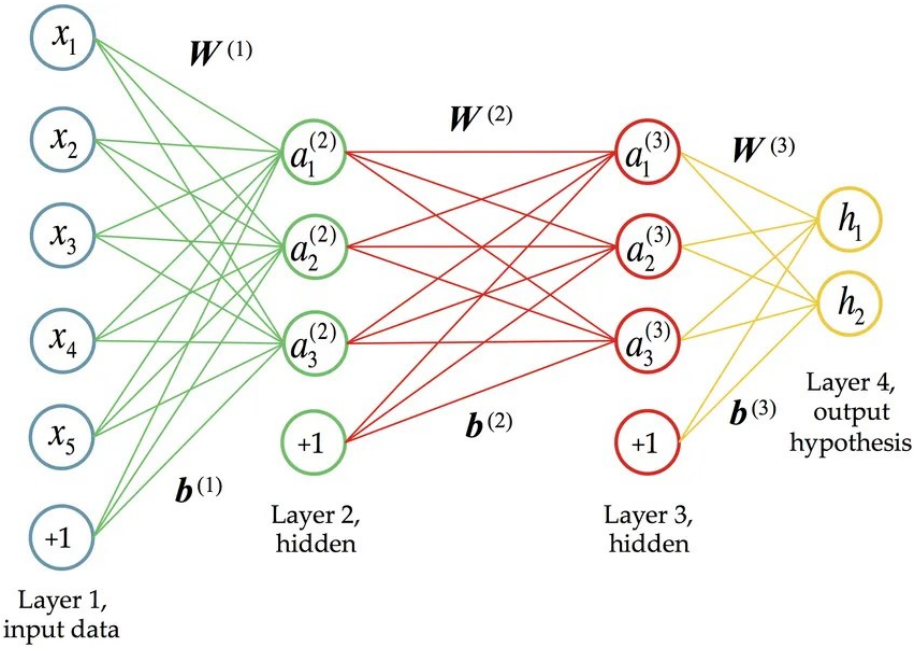

Figure 2. Example of neural network structure architecture with two hidden layers (Schmidhuber 2015).

This study adopts a supervised learning approach, which requires labelled data during training; therefore, hyperparameters must be specified before the training process (Probst et al., 2018). The selection of hyperparameters is particularly important in predicting elastic properties because subsurface elastic responses are governed by complex and nonlinear interactions among lithology, porosity, fluid content, and pressure conditions. Machine-learning models are capable of capturing these nonlinear relationships and integrating multiple petrophysical logs simultaneously, offering advantages over conventional linear or empirical methods.

In this research, hyperparameters are categorized into four main categories: (1) Architectural hyperparameters, including the number of hidden layers, numbers of neurons per layer, activation functions, and output functions; (2) Learning hyperparameters, such as learning rate, optimizer, batch size, numbers of training epochs, and weight initialization; (3) Regularization hyperparameters, involving dropout rate, batch normalization, and early stopping, which applied to prevent overfitting; and (4) Data and training strategy hyperparameters, comprising validation split ratio, data shuffling, scaling method, loss function, and learning-rate scheduling. The dataset is divided into training and test sets, with the training data used to build the machine-learning models. Hyperparameter tuning and cross-validation are subsequently performed to ensure that the model produces consistent predictions (Wardhana 2021).

Each hyperparameter is carefully selected to guarantee that the model accurately captures the nonlinear behaviour of elastic parameters while maintaining stability and generalization, especially under data-scarce conditions. In this study, Mean Squared Error (MSE) is used as the loss function (see Equation 4), as it penalizes large prediction errors and is well suited for regression-based elastic parameter estimation. The application of MSE facilitates optimization of the FCNN model for precise prediction of Vs and Vp/Vs logs, thereby enhancing the reliability of elastic property characterisation.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(yi - \tilde{y}i)^2 \quad (4)$$

where N is the number of samples, y_i represents the actual (true) value, and $\hat{y}_i$ represents the predicted value. The MSE quantifies the average squared difference between the predicted and actual values; smaller MSE values represent better model performance.

Data preparation

This study utilizes data from the East Java Field, consisting of two wells: one with available Vs log data (Well 2) and one without Vs measurements (Well 1). Due to these data limitations, the analysis emphasizes prediction of the Vp/Vs log rather than direct estimation of Vs in all intervals.

The two wells also differ in penetration depth. The shallower well contains measured Vs data, whereas the deeper well does not. This discrepancy may affect machine learning-performance, as the training dataset is dominated by shallow interval information. To mitigate this limitation, data from both wells are integrated into the training process. Additionally, the Castagna equation is applied as a reference constraint to guide predictions in Well 1.

Prior to model development, feature selection is conducted to determine which logs should be used as input variables. A correlation matrix is generated to identify the logs most strongly correlated with the target variable, as shown in Figure 3. This matrix illustrates the degree of linear association among logs, allowing selection of the most relevant for training the machine-learning model. Based on the result above, the most relevant features are selected while non-informative features are removed. This process improves model performance by reducing data complexity and minimizing the risk of overfitting (Zainuri et al., 2023). The correlation between Vp/Vs and Vs is presented in the Table 1 below.

Table 1 indicates that only *in-situ* well-logs – directly measured within the borehole – are used, ensuring that the data accurately represent actual subsurface condition. The table demonstrates the relationship between the in-situ logs and the predicted Vs and Vp/Vs logs, demonstrating a reasonably strong correlation. The results suggest that Vp/Vs prediction is sufficiently reliable, as the overall trend is consistently preserved when compared with the original well log data.

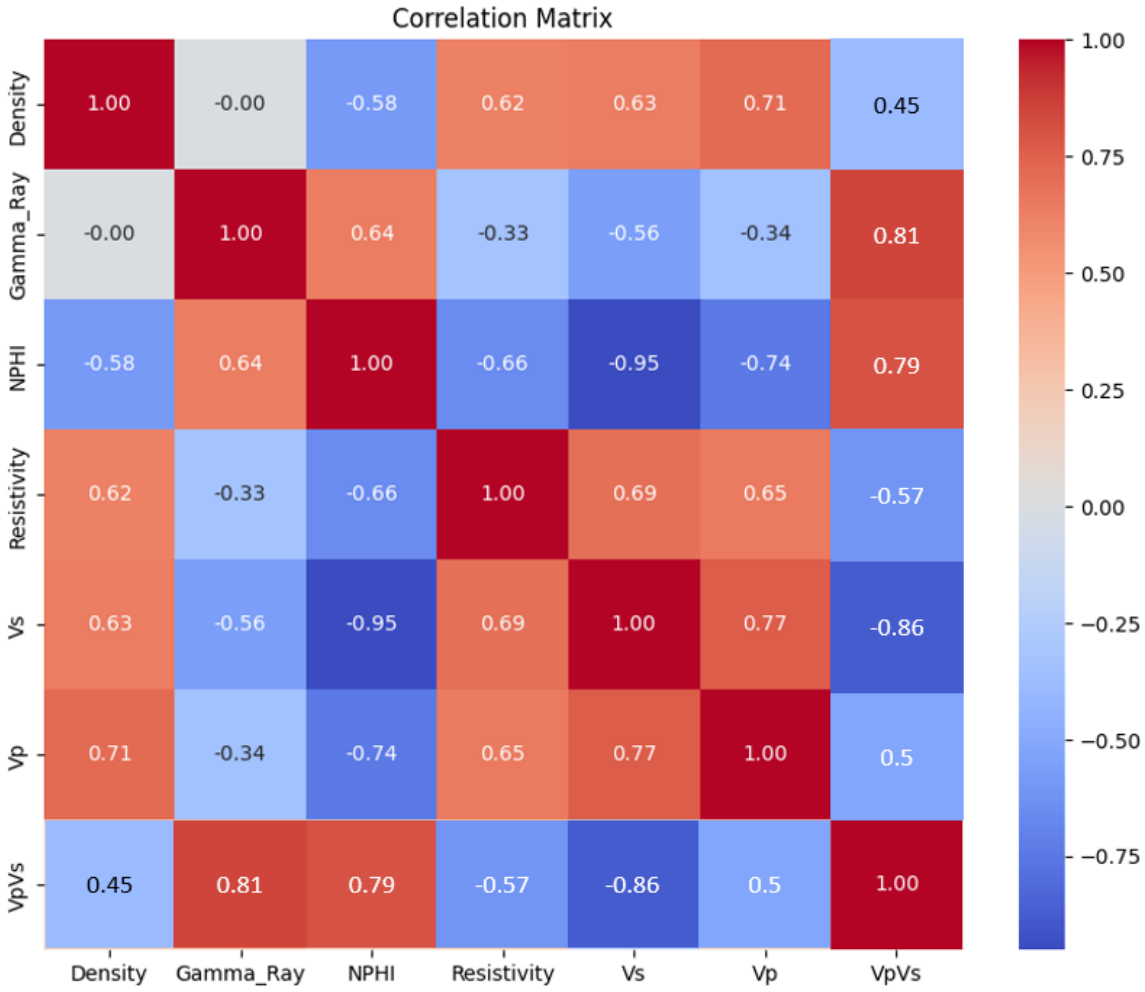

Figure 3. Correlation matrix showing relationships among well logs.

In this study, Vs obtained from Multiple Linear Regression (MLR) is used to support machine-learning training process. Conventionally, Vp logs are used as a reference predicting Vs because Vp and Vs are known to exhibit a predominantly linear relationship. However, Vp is highly sensitive to fluid content, whereas Vs is relatively insensitive to fluid variations. Consequently, if the predicted Vs exhibits fluid sensitivity similar to Vp, it may not accurately represent the true shear-wave velocity. To address this issue, Vs derived from the multi-attribute MLR approach is incorporated as training input for machine learning model.

Once all relevant features are selected, the machine-learning model can be constructed. The initial step involves defining the appropriate hyperparameters for model development. Careful hyperparameter selection is essential to prevent overfitting and ensure that the trained model can generalize effectively to other wells.

Data normalization

Data scaling is an essential preprocessing step of machine-learning model development, especially for gradient-based optimization algorithms such as those used in Fully Connected Neural Networks (FCNN) (Goodfellow, 2016). The primary objective of scaling is to ensure that all input features share a comparable numerical range, enabling stable and efficient model training.

Without proper scaling, features with large numerical values (e.g., the Sonic log measured in microseconds per foot) may dominate the weight update process, while features with smaller numerical ranges (such as porosity or gamma-ray

Table 1. Result of Vs and Vp/Vs correlation to other logs.

Attribute	Correlation with Vs	Correlation with Vp/Vs
Vp	0.77	0.5
Vs	1	-0.86
Resistivity	0.69	-0.57
Gamma Ray	-0.56	0.81
Neutron Porosity	-0.95	0.79
Density	0.63	0.45

logs) may contribute less significantly. This imbalance can lead to slow convergence, unstable training, or even failure to converge due to disproportionate gradients updates. In this study, the Robust Scaler is employed as the normalization technique for all logs used in model training. The Robust Scaler utilizes median and interquartile range (IQR) statistics for scaling, making it particularly suitable for datasets containing outliers. As illustrated in Function 5, the Robust Scaler transforms the data using quartiles and IQR, thereby minimizing the influence of extreme values.

Well-log data often exhibit significant variability caused by lithological heterogeneity or fluid presence. Therefore, the Robust Scaler is considered appropriate for this application, as it preserves important geological variations while minimizing distortion from outliners. The equation for the Robust Scaler is presented in Equations 5 and 6.

$$x^{'} = \frac{x - median\ (x)}{IQR(x)} \quad (5)$$

$$IQR\ (x) = Q3 - Q1 \quad (6)$$

The original data value (such as a log measurement like density, gamma ray, or Vs) is represented by x. The dataset's median, or median (x), is the middle value when the data are sorted. It is a reliable indicator of central tendency that is less susceptible to outliers than the mean. The interquartile range, represented by the quantity IQR (x), is the difference between the first quartile (Q1) and the third quartile (Q3), where Q1 is the dataset's 25th percentile and Q3 is its 75th. By deducting the median from the initial value and dividing by the IQR, the normalized value x′ (x prime) is produced. This successfully scales the data according to its spread while reducing the impact of value.

Model evaluation

In the model evaluation stage, the selected hyperparameters utilized in developing the machine-learning models are clearly defined. A key consideration is the proportion of data allocated for training and validation. This study examines the impact of different training–validation split ratios to assess the model's generalization capability, particularly under limited data conditions.

The choice of activation function and the number of hidden layers form part of the network architecture design. A four-layer neural network is implemented in this study to balance model complexity and overfitting risk while preserving the intrinsic structure of the data. The Rectified Linear Unit (ReLU) activation function is adopted due to its computational efficiency and ability to facilitate stable gradient propagation, making it well suited for regression-based modeling task.

Model evaluation metrics and regularization parameters are vital for performance assessment. In this study, mean absolute error (MAE) is used as the primary evaluation metric to quantify prediction accuracy. In addition, the Pearson correlation coefficient is applied to measure the strength and direction of the linear relationship between predicted and actual values. Equations 7 and 8 present the mathematical formulation of MAE and the Pearson correlation coefficient, respectively.

$$MAE = \frac{1}{x}\sum_{i=1}^{n}|m_i - n_i| \ldots\ldots \quad (7)$$

$$r = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\dot{y}_i - \bar{\dot{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\dot{y}_i - \bar{\dot{y}})^2}} \ldots\ldots \quad (8)$$

For equation 7, m_i represents the measured value, n_i denotes the predicted value, and x is the total number of samples. For equation 8, y_i is the actual value at i-th data point, $\dot{y}_i$ is the corresponding predicted value, $\bar{y}$ is the mean of the actual values $\hat{y}$ is the mean of the predicted values, and N is the total number of observations.

A well-performing model is characterized by a Pearson correlation coefficient approaching 1 and a low MAE value, indicating minimal prediction error and strong agreement between predicted and observed data.

Figure 4 illustrates the workflow for predicting the Vp/Vs log using a Fully Connected Neural Network (FCNN). Several factors are considered during model construction, including the learning rate, the number of hidden layers, and the number of neurons in each layer. These parameters significantly influence model accuracy, convergence behavior, and generalization performance, while also helping to reduce overfitting. In this study, a learning rate of 0.001 is employed to ensure stable convergence and alignment between predicted outputs and the distribution of actual log data. Additionally, regularization techniques such as early stopping and batch normalization are implemented during model training to further mitigate overfitting and enhance generalization across different wells.

After identifying the optimal model configuration, the trained FCNN is applied to predict the Vp/Vs log in Well-2 , which contains complete measured data. The predicted Vp/Vs values are then compared with the corresponding actual measurements to evaluate model performance. If the resulting correlation and error metrics are satisfactory, the trained model is subsequently applied to predict the Vp/Vs log for Well-1.

RESULT AND DISCUSSION

This study addresses several key aspects necessary for accurate elastic property prediction. As discussed previously, the Vs log is predicted using both Multiple Linear Regression (MLR) and the Castagna equation. However, within the machine-learning workflow, the primary objective is not to directly predict Vs, but rather to estimate the Vp/Vs ratio. This methodology leverages the capability of machine-learning models to capture non-linear relationships between well-log attributes and the Vp/Vs log.

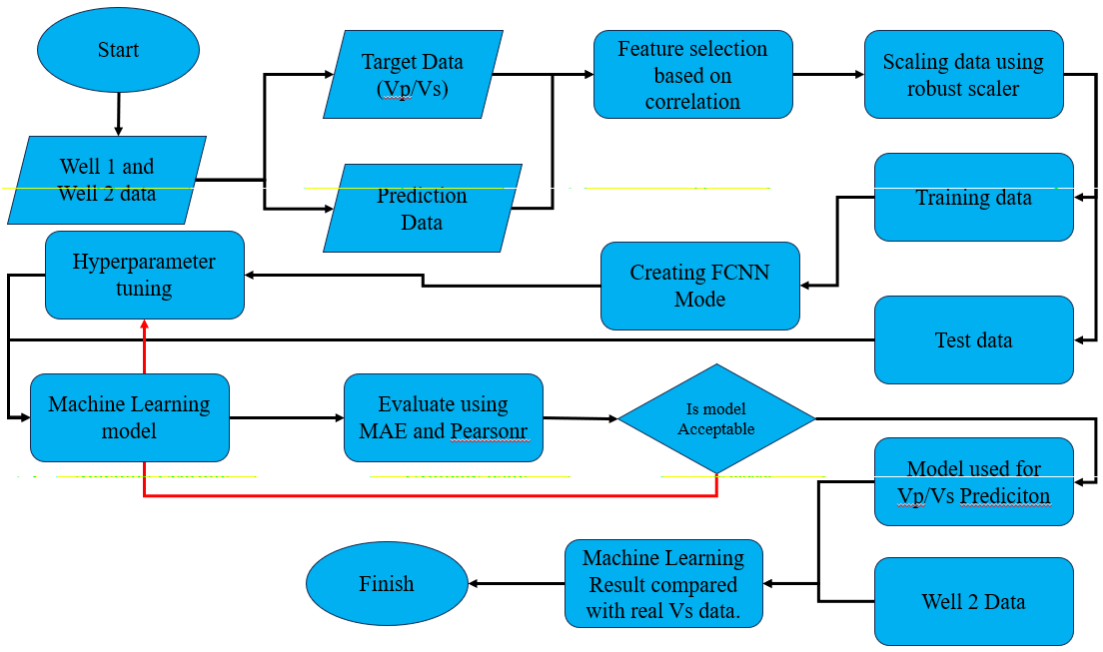

Figure 4. Workflow for building machine-learning model

The develop the machine learning model begins with defining appropriate hyperparameters and validation scenarios. Ulil Albab (2025) demonstrated that hyperparameter selection significantly influences model performance. Rather than extensively modifying architectural hyperparameter, this study focuses on varying the proportions of training and validation datasets to investigate their impact on predictive accuracy.

The training dataset is selected based on strong correlations observed in Figure 3. The input logs comprise NPHI, GR, RHOB, Resistivity, and Vs derived from the MLR approach. Multiple modeling scenarios are designed to assess model performance under different data-splitting configurations, as detailed in Table 2. Using a random split method across both wells, the training proportion is varied from 50% to 70%. The machine-learning model exhibits robust performance under the 70% training configuration. When comparing the Vs log derived from the predicted Vp/Vs ratio with the actual Vs log from Well-2, the model achieves a correlation coefficient of 0.8204.

Further analysis shows that even when the training subset is reduced relative to the validation set, the model maintains a high correlation, demonstrating the robustness of machine-learning approach. Consequently, the trained model is considered reliable for predicting the Vp/Vs log in Well-1. Subsequently, Vs logs obtained from MLR, the Castagna equation, and the machine-learning model are compared using Well-2 data, which contains measured Vs values.

Based on the correlation results in Table 3, MLR produces higher correlation values compared to other methods due to its ability to identify linear relationships among log data. However, its performance differs when fluid effects are present. As shown in Figure 8, both MLR and Castagna methods fail to clearly identify gas-bearing intervals.

Although Figure 5 indicates that all methods show strong correlation with the measure Vs log in Well-2 and that differences among methods appear minimal, this similarity in the Vs domain does not necessarily imply equivalent performance in reservoir interpretation. All methods appear reliable when evaluated solely against Vs measurement. However, a critical consideration is the role of Vp. According to the Castagna relationship, Vs is linearly related to Vp, yet Vp is significantly more sensitive to fluid content than Vs.

To minimize fluid-drive bias, this study predicts Vs indirectly through direct estimation of Vp/Vs without explicitly using Vp as an input in the machine-learning process. By directly predicting the Vp/Vs, the model learns nonlinear interactions among the input logs and generates Vs estimates

Table 2. Validation data scenarios used in the machine-learning model.

No	Data validation/test	Parameter					Correlation of Vs
		Hidden layer	Batch size	Epoch	Activation function	Learning rate	Well 2
1	70 %/30 %	4	128	1000	ReLu	0.001	0.8204
2	60 %/40 %	4	128	1000	ReLu	0.001	0.8236
3	50 %/50 %	4	128	1000	ReLu	0.001	0.8263

Table 3. Comparison of Vs predictions from all methods.

No	Method	Correlation
1	Machine learning	0.8204
2	Multi-linear regression	0.9384
3	Castagna wet limestone	0.7801

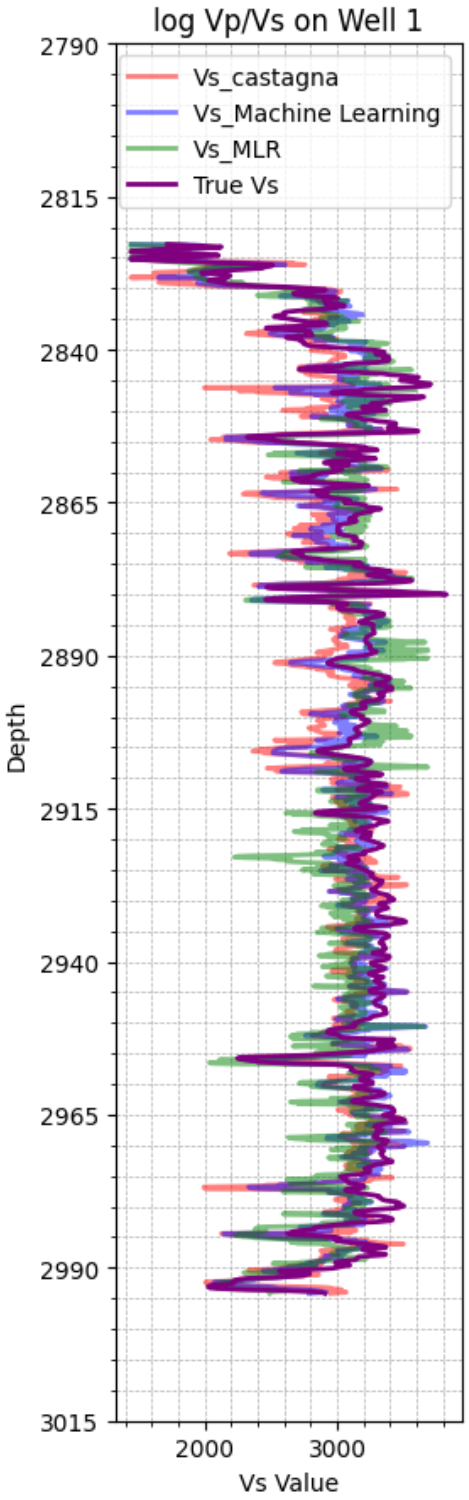

Figure 5. Comparison of elastic properties prediction from all methods (Well-2). With red line is Vs_castagna, Blue line is Vs machine learning, green line is Vs MLR and purple line is real Vs data.

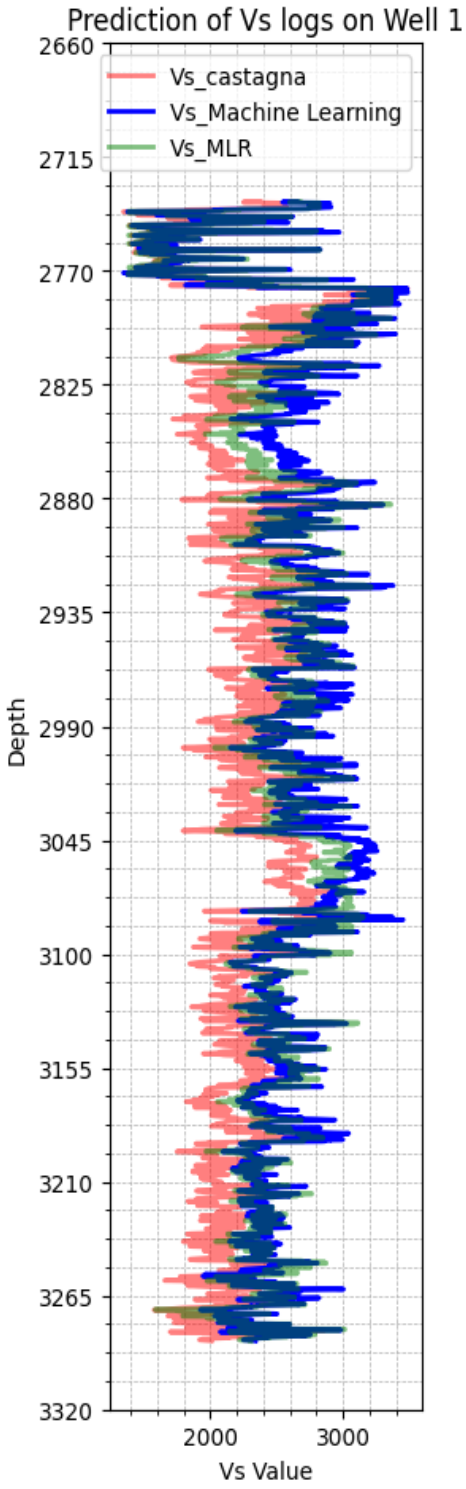

Figure 6. Comparison of Vs Prediction from all methods in Well-1. With red line is Vs_castagna, Blue line is Vs machine learning and green line is Vs MLR.

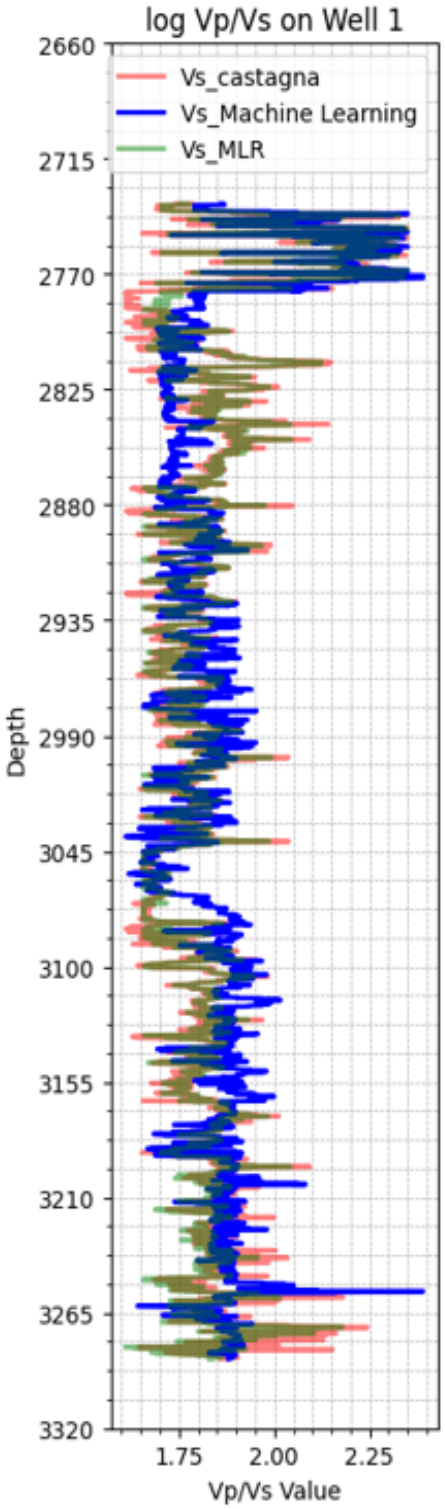

Figure 7. Comparison of Vp/Vs logs from all methods in Well-1 . With red line is Vs_castagna, Blue line is Vs machine learning and green line is Vs MLR

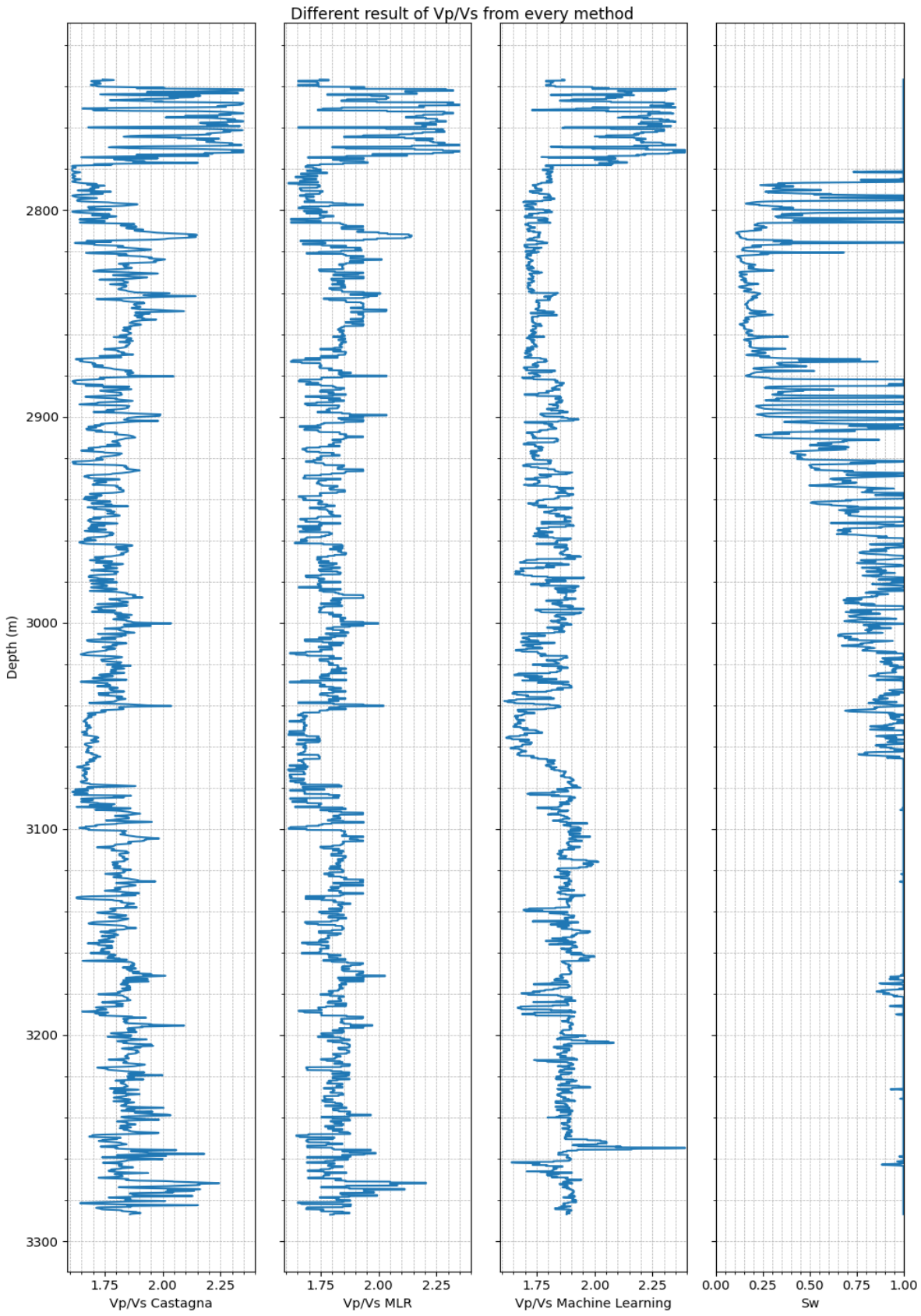

Figure 8. Comparison of Vp/Vs logs from all methods in Well-1 with the Sw log.

that are less influenced by fluid sensitivity. Machine-learning techniques are particularly suitable for this task due to their ability to model complex nonlinear dependencies.

After achieving satisfactory results in Well-2, evidenced by strong agreement between predicted and measured Vs values, the trained model is applied to Well-1. Figure 6 presents the predicted Vs logs. The red curve represents the Castagna-derived Vs, the green curve represents MLR-derived Vs, and the blue curve represents the machine-learning predictions. While the overall trends appear similar, the Castagna method consistently produces slightly lower Vs values compared to the other approaches.

However, when these Vs results are transformed into Vp/Vs ratios – an essential parameter for reservoir characterization – clear differences emerge. Although Vs logs appear comparable, converting them to Vp/Vs highlights more pronounced discrepancies among the methods. This emphasizes the importance of evaluating elastic property predictions not only in the Vs domain but also through derived elastic parameters such as Vp/Vs, which provide more direct insight into subsurface fluid conditions.

Figure 7 shows the Vp/Vs ratios obtained from each method. In conventional approach, Vp is used to derive Vp/Vs, whereas the machine-learning model predicts Vp/Vs directly. The analysis focuses on the red-highlighted. The Castagna derived Vp/Vs (red curve), and MLR-derived Vp/Vs (green curve) predominantly indicate water-saturated conditions. However, the Sw log clearly shows reduced water saturation at specific depths, suggesting the presence of gas.

Under gas saturation conditions, Vp/Vs is expected to decrease due to a reduction in Vp while Vs remains relatively stable. Neither Castagna nor MLR effectively captures this behavior. In contrast, the machine-learning prediction (blue curve) accurately reflects the gas-bearing interval by detecting the reduction in Vp/Vs corresponding to the Sw anomaly. This demonstrates that direct prediction of Vp/Vs allows the model to capture subtle nonlinear relationship influenced by fluid presence, which are not adequately represented by traditional empirical methods.

These findings indicate that elastics property prediction is more effective when performed directly on parameters used to reservoir characterization, such as Vp/Vs, Poisson's ratio, or lambda–mu–rho attributes. Although Vs is valuable for lithological interpretation due to its relative insensitivity to fluid, predicting Vs using fluid-sensitive inputs may introduce ambiguity.

A more robust approach involves directly predicting fluid-sensitive elastic attributes such as Vp/Vs while subsequently deriving Vs if required. This approach reduces uncertainty, improves of fluid interpretations, and yields elastic responses that more accurately reflect actual subsurface conditions.

CONCLUSION

The results of direct prediction of Vp/Vs logs demonstrate strong performance, aligning with expectations for reservoir characterization. The predicted Vp/Vs ratios effectively identify fluid variations, highlighting the capability of machine-learning techniques to model complex subsurface relationships.

This study also evaluated various training-validation configurations and finds that even limited training data can produce satisfactory predictive results. Comparison between machine-learning predictions and measured data in Well-2 yields a correlation coefficient of approximately 0.8204, indicating the robustness and reliability of FCNN approach for elastic property prediction.

Overall, direct production of Vp/Vs proves more effective than indirect Vs-based approached for fluid-sensitive reservoir analysis . This study show that machine learning able to predict the condition of Vs log with the presence of fluids by utilizing Vp/Vs approach. This study show the limitation of depth has an effect to machine learning initial build model that could cause miss interpretation on the result of prediction so that's why by using Castagna method to help the learning process for the machine learning in the missing Vs.

ACKNOWLEDGEMENT

The author sincerely acknowledges the Bandung Institute of Technology (ITB) for providing academic support, research facilities and data access that contributed to the successful completion of this study.

GLOSSARY OF TERMS & SYMBOLS

Terms & Symbols	Definition	Unit
Vp	Velocity of P-wave	m/s
Vs	Velocity of S-wave	m/s
RHOB	Rock's density	g/cc
NPHI	Neutron porosity	%
GR	Gamma ray	API
Sw	Water saturation	0.Sw
FCNN	Fully connected neural network	
ANN	Artificial neural network	
MLR	Multiple linear regression	
MAE	Mean absolute error	

REFERENCES

Akhundi, H., Ghafoori, M. and Lashkaripour, G. ,2014, Prediction of Shear Wave Velocity Using Artificial Neural Network Technique, Multiple Regression and Petrophysical Data: A Case Study in Asmari Reservoir (SW Iran). Open Journal of Geology, 4,303-313. https://doi.org/10.4236/ojg.2014.47023

Alfarraj, M., & AlRegib, G. (2021). Deep learning for elastic property prediction from seismic data: A review. Geophysics, 86(6), W1–W15. https://doi.org/10.1190/geo2020-0835.1

Candra, A. D., Rahalintar, P., Sulistiyono, & Prabowo, U. N. (2024). Comparison of facies estimation of well log data using machine learning. Scientific Contributions Oil and Gas, 47(1), 21–30. https://doi.org/10.29017/SCOG.47.1.1593

Castagna, J. P., Batzle, M., Eastwood, R. L., Johnston, D. H., & Meissner, R. (1993). Velocity–porosity relationships for siliciclastic and carbonate sediments from sonic and density logs. Geophysical Prospecting, 41(2), 149–187.

Feng, G., et al. (2023). Shear-wave velocity prediction based on deep neural networks and rock physics. Frontiers in Earth Science, 11.

Fu, R., Zhang, Z. & Li, L., 2016, Using LSTM and GRU neural network methods for traffic flow prediction, Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 324-328.

Fu, X., Wei, Y., Su, Y. & Hu, H., 2024, Shear wave velocity prediction based on the long short -term memory network with attention mechanism, Applied Sciences, vol. 14, pp. 2489. https://doi.org/10.3390/app14062489

Kim, Sujeong., et al. (2025), Machine Learning-Based Prediction of Subsurface Elastic Properties Using Seismic and Well-Log Data, EEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing18:1-20. DOI: 10.1109/JSTARS.2025.3623096

Liu, H., Grana, D., & Mukerji, T. (2024). Physics-informed machine learning for elastic property estimation in reservoir characterization. IEEE Transactions on Geoscience and Remote Sensing, 62, 1–13. https://doi.org/10.1109/TGRS.2023.3321457

Lishner, I. & Shtub, A., 2022, Using an artificial neural network for improving the prediction of project duration, Mathematics, vol. 10,pp.4189. https://doi.org/10.3390/math10224189

Mahmoud, A. A., Elkatatny, S., & Abdulraheem, A. (2022). Prediction of compressional and shear wave velocities using machine learning techniques. Journal of Petroleum Science and Engineering, 208, 109336. https://doi.org/10.1016/j.petrol.2021.109336

Mehrad, M., et al. (2022). Estimating shear wave velocity in carbonate reservoirs using machine learning. Journal of Petroleum Science and Engineering.

Montgomery, Douglas C, et al., 2012, Introduction to Linear Regression Analysis Fifth Edition Schmidhuber, J. (2015). Deep learning in neural

networks: An overview. Neural Networks, 61, 85–117.

Ulil, M. R., Winardhi, S., & Dinanto, E. (2025). Machine Learning-Based Prediction of Shear Wave Velocity: Performance Evaluation of Bi-GRU, ANN, and The Greenberg-Castagna Empirical Method. Scientific Contributions Oil and Gas, 48(3), 159-171. https://doi.org/10.29017/scog.v48i3.1797

Rajabi, M., et al. (2022). Predicting shear wave velocity from conventional well logs with deep and hybrid machine learning algorithms. Journal of Petroleum Science and Engineering.

Sudrazat, S. D., Purba, H., Wijaksono, E., Pranowo, W., & Hibatullah, M. I. (2020). Prediksi kecepatan gelombang S dengan machine learning pada sumur "S-1", Cekungan Sumatera Tengah, Indonesia. Lembaran Publikasi Minyak dan Gas Bumi, 54(1), 29–35. https://doi.org/10.29017/LPMGB.54.1.502

Suleymanov, V., El-Husseiny, A., Glatz, G., & Dvorkin, J. (2023). Rock physics and machine learning comparison: Elastic properties prediction and scale dependency. Frontiers in Earth Science.

Xu, Shiyu & Payne, Michael, 2009, Modelling elastic properties in carbonate rocks

Zhang, Y., et al. (2022). Robust estimation of shear wave velocity in a carbonate oil reservoir using machine learning algorithms. Energy Reports.

Zainuri, A. P. P., Sinurat, P. D., Irawan, D., & Sasongko, H. (2023). Trap prevention in machine learning in prediction of petrophysical parameters: A case study in the Field X. Scientific Contributions Oil and Gas, 46(3), 115 –127. https://doi.org/10.29017/SCOG.46.3.1586