

The Effectiveness of Integrating Machine Learning Approach with Near-Surface Seismic Data for Lithologic Zones Determination in the Krueng Kaleng Bridge Area, Lamno, Aceh Jaya

^{1*}Tika Hapsari, ²Khaizal Jamaluddin, ¹Dina Gunarsih, ¹Nazla Syafitri Rangkuti, ³Tarmizi, ³Uswatun Hasanah

¹Program Studi Teknik Geologi, Universitas Syiah Kuala, Kota Banda Aceh, Indonesia

²Program Studi Teknik Sipil, Universitas Syiah Kuala, Kota Banda Aceh, Indonesia

³Program Studi Teknik Geofisika, Universitas Syiah Kuala, Kota Banda Aceh, Indonesia

*Corresponding: tikahapsari@usk.ac.id

Manuscript received: August 1, 2025; Revised: August 12, 2025; Accepted: August 16, 2025

ABSTRACT

A scientific approach to determine the lithological zone of the Krueng Kaleng bridge construction area, Lamno, Aceh Jaya is carried out by integrating a machine learning approach to geophysical methods. These lithologic zones need to be known in order to provide recommendations a depth of foundations that can support the bridge. This machine learning processing uses input data V_p and V_s generated from geophysical methods. Lithological zone determination can also be done by analyzing statistics using PCA and K-Means from variations in physical parameter data. Borehole data is also used as secondary data to validate the final results of lithologic zone interpretation. Based on the results of data processing and interpretation, 3 layers were obtained in the study area with V_p ranging from 0.31 - >2 km/s and V_s ranging from 60 - 400 m/s. The lithological zone consists of surface soil layer, saturated clay sand layer, and unsaturated clay sand layer to a depth of more than 13 meters. The results of machine learning processing using Google Colab are considered effective because they produce the same number of zones, namely three zones. This interpretation can also be validated with borehole data described as containing clay and sand. Based on the geological map, the study area is located on volcanic rocks which can be interpreted that this near-surface layer is likely to experience high alteration and weathering resulting in a secondary lithology of clay. Therefore, it is not recommended to build foundations less than 20 meters.

Keywords: Near-Surface Seismic, V_p/V_s Ratio, Sedimentology, Machine Learning, Lamno.

ABSTRAK

Pendekatan ilmiah untuk menentukan zona litologi pada area pembangunan jembatan Krueng Kaleng, Lamno, Aceh Jaya dilakukan dengan mengintegrasikan pendekatan machine learning dengan metode geofisika. Zona litologi ini perlu diketahui agar dapat memberikan rekomendasi kedalaman fondasi yang dapat menopang jembatan. Pengolahan machine learning ini menggunakan data input V_p dan V_s yang dihasilkan dari metode geofisika. Determinasi zona litologi juga dapat dilakukan dengan menganalisis statistik menggunakan PCA dan K-Means dari variasi data parameter fisis. Data lubang bor juga digunakan sebagai data sekunder untuk memvalidasi hasil akhir interpretasi zona litologi. Berdasarkan hasil pengolahan dan interpretasi data, didapatkan 3 lapisan pada daerah penelitian dengan V_p berkisar antara 0,31 - >2 km/s dan V_s berkisar antara 60 - 400 m/s. Zona litologi terdiri dari lapisan tanah

permukaan, lapisan pasir lempung jenuh, dan lapisan pasir lempung tidak jenuh hingga kedalaman lebih dari 13 meter. Hasil pengolahan machine learning dengan menggunakan Google Colab dinilai efektif karena menghasilkan jumlah zona yang sama, yaitu tiga zona. Interpretasi ini juga dapat divalidasi dengan data lubang bor yang dideskripsikan mengandung lempung dan pasir. Berdasarkan peta geologi, daerah penelitian terletak pada batuan vulkanik yang dapat diinterpretasikan bahwa lapisan dekat permukaan ini kemungkinan mengalami alterasi dan pelapukan yang tinggi sehingga menghasilkan litologi sekunder lempung. Oleh karena itu, tidak direkomendasikan untuk bangun pondasi kurang dari 20 meter.

Kata Kunci: Seismik Permukaan, Rasio Vp/Vs, Sedimentologi, *Machine Learning*, Lamno

1. INTRODUCTION

The Lamno region can be found in Aceh Jaya district and is surrounded by diverse landscapes, including mountains, hills, lowlands, and coastlines. These physiographic conditions can be distinguished from one another by observing the visible surface conditions. The current state of the study area is mountainous and covered by forest and shrubs. This area with topographic elevation reaching 200 – 400 m above sea level. This area is the construction area of the Jantho to Lamno through road, which is located on the edge of the river and is planned for the construction of a bridge connecting the central crossing road. This area of the bridge is referred to as Krueng Kaleng bridge in the Sabet village. Understanding the physical properties of soil is crucial before building a bridge. This data on subsurface soil physical properties is essential for professionals and planners as a foundation for evaluation and deliberation before making decisions in the design of a bridge. Assembling subsurface information requires data collection through geophysical methods.

This study emphasizes the combination of machine learning approach with geophysical methods, specifically refraction seismic and Multichannel Analysis Surface Waves (MASW) methods, to acquire subsurface physical properties that will subsequently be corroborated with additional

data from geological maps and drilling data at a 5-meter depth, aiming to identify lithological zones within the research area.

The geology of Sabet Village, Aceh Jaya mainly lies within the Bentaro Volcano rock formation (Muvb). This formation consists of basalt rocks, agglomerates, basalt formation, and mafic fractures. This formation is one of the surface layers within the Woyla Group located in Aceh, resulting from graben formation through faults, and is distributed in two sections toward the northeast and southwest of the Sumatra Fault. The Woyla Group unit is commonly located in the Barisan Mountains. According to the geological map [1], there are two uncertain faults situated near the study area. In the northern section of the measurement zone lies the Lhoong rock formation, which is an associated formation and predates the Gunungapi Bentaro formation. In the western region of the zone, the Lamno Limestone Formation was identified, which is connected to the exposure of the Bentaro Volcano Formation, consisting of dark limestone and volcanic rock debris. The Bentaro Volcano Formation originated in the Jurassic Period, making it approximately 200 million years old. The geologic map of this study area with the bore data point shown in Figure 1.

To classify the lithology zones of the study area, the authors utilized clustering techniques like Principal Component Analysis (PCA) and K-Means implemented

through machine learning, subsequently combined with the outcomes of refraction and MASW seismic data processing, which were later corroborated with drill data and geological maps to be discussed in section 3. The objective of identifying the lithological zone is to understand the subsurface conditions, such as the presence of loose or hard layers, which will affect the construction of the Krueng Kaleng bridge and provide recommendations regarding the depth of the foundation. Thus, this integration method aims to explore how machine learning clustering can be effectively connected to traditional geophysical methods in its interpretation. Using machine learning tools can process data more quickly and accurately, but there still has to be data validation with subsurface imaging.

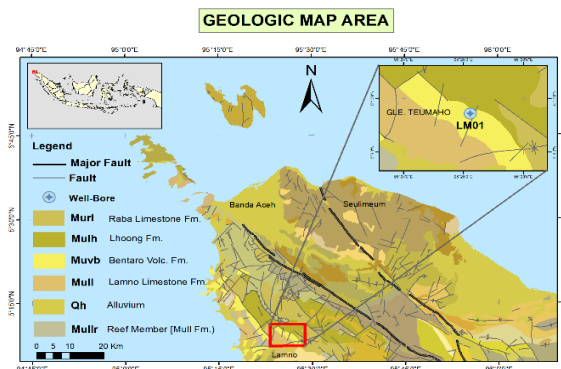


Figure 1. Geological map research area, with LM01 bore data point.

The subsurface engineering image is complex and challenging to interpret manually, with reference values of physical parameters having a wide range of possible errors [2]. Interpretation of subsurface images requires a deep understanding of physical properties, including geology processes and rock mechanics. Physical properties such as V_p/V_s ratio, elastic modulus, poisson ratio, and density. Physical properties refer to the characteristics of the medium that seismic waves travel through, with each layer of rock

exhibiting different seismic wave velocity values. Refraction seismic employs the concept of seismic wave refraction (P-waves (V_p)) as they travel through rock layers, where seismic MASW emphasizes examining surface waves like Rayleigh waves, which respond significantly to variations in the shear velocity (V_s) of the rock layers. The equations for the P-wave and S-wave velocities are as follows.

$$V_p = \left(\frac{\lambda + 2\mu}{\rho} \right)^{1/2} = \left(\frac{k + \frac{4}{3}\mu}{\rho} \right)^{1/2} \quad (1)$$

$$V_s = \left(\frac{\mu}{\rho} \right)^{1/2} \quad (2)$$

V_p means P-wave velocity and V_s means S-wave velocity, where λ is the lame parameter, μ is the rigidity modulus (N/m^2), k is the bulk modulus (GN/m^2) and ρ is density (kg/m^3). The parameters such as k and μ refer to the modulus of elasticity and density are directly related to the ability of the medium to receive the force (noise) in the form of seismic waves. These parameters can be determined from seismic wave velocity values to determine the type of rock layer based on equations (1) and (2).

The calculation results according to this equation are obtained as input data for PCA and K-Means analysis performed on machine learning tools using Python coding. The main concept of PCA as a statistical method is for reducing the dimensionality of a data set containing numerous irrelevant variables, while retaining as much of the variation in the data set as possible so that the real information can still be preserved [3][4]. Then, K-Means is also a statistical method used to classify objects by dividing the data into groups based on similar characteristics in each group according to the closest distance [5]. Numerous researchers have examined the

use of cluster analysis using seismicity data. Several topics use this cluster analysis, such as used to determine geological strike [6], for channel delineation, reservoir characteristics [7], and earthquake clustering [8].

2. MATERIAL AND METHOD

Seismic data recordings were carried out on a single refraction seismic line and one aligned MASW seismic line. The path is aligned NW-SE and is adjacent to the Krueng Kaleng river. The instruments utilized for field data collection include the PASI 16S24-P seismograph and associated equipment, sledgehammer, metal plate, 12 V battery, GPS, geological compass, and measuring device. Data processing utilizes various software including ZondST2D, Seis-Imager, Surfer, and Google Colab. Refraction seismic data is analyzed with ZondST2D software, whereas MASW seismic data is analyzed using Seis-Imager software. Data interpretation was conducted by looking at the reference values of V_p and V_s , then calculating the two velocities into V_p/V_s in Surfer software. The physical results of V_p/V_s were later calculated to obtain the poisson ratio and elastic modulus. P-wave velocity is obtained from the results of processing refraction seismic data using ZondST2D and S-wave velocity is obtained from results of processing Rayleigh surface waves on MASW seismic data using SeisImager that utilize the dispersion properties of surface waves. The reference values for V_p , V_s , and V_p/V_s are as **Table 1**, **Table 2**, and **Table 3**.

Table 1. Reference values of P-wave velocity for materials [9]

Material	V_s (km/s)
Top soil	0.4 – 1.0
Saturated clayey sand	1.0 – 1.7
Unsaturated clayey sand	1.7 – 2.1
Weathered rock	2.1 – 2.7
Sandstone	>2.7

Table 2. Reference values of S-wave velocity for soil type [10]

Soil Type	V_s (m/s)
SA (hard rock)	>1500
SB (rock)	750 to 1500
SC (hard soil, very dense, and soft rock)	350 – 500
SD (medium soil)	175 to 350
SE (soft soil)	<175
SF (special soil)	Requires specific investigations

Table 3. Reference values of V_p/V_s Ratio [11]

Type of Rock	V_p/V_s Ratio
Fine grained sand	0.0 – 1.2
Medium-grained sand	1.2 – 1.45
Coarse-grained sand	1.46 – 1.6
Sandstone	1.6 – 1.8
Shale or Clay	Above 2.0

Based on these reference values, it is possible to interpret the lithology type in the study area. The results of this geophysical method are then compared with cluster analysis using the Google Colab machine learning tool. This Google Colab platform can be used in Python programming language that can run PCA and K-Means analysis [12]. The data inputted on the platform are physical values such as V_p/V_s , Poisson Ratio, Modulus of Elasticity (E), and density values which are then classified into several types of lithology. The flow chart of this research is as follows in Figure 2.

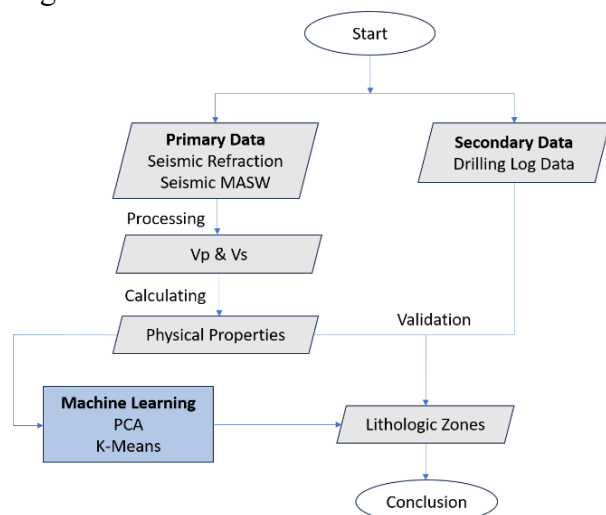


Figure 2. Research flow chart

3. RESULT AND DISCUSSION

3.1 P-Wave Velocity

Interpretation of the refraction seismic data processing resulted in wave velocities ranging from 0.31 km/s to >2 km/s. The subsurface image of the refraction seismic data processing results as shown in Figure 3 on the color bar scale indicates that the reddest color indicates a smaller V_p value than the bluest color. Based on the 2D cross section profile, it can be concluded that it is divided into 3 rock layers [11], namely at a thickness of 1 to 3 meters containing saturated clayey sand which ranges from 1 - 1.7 km/s which is named cluster 0, with a thickness of 3 meters including the top soil layer which ranges from 0.31 - 1 km/s which is named cluster 1, and with a thickness of more than 3 meters there is a layer of unsaturated clayey sand which ranges from 1.7 - 2.1 km/s which is named cluster 2.

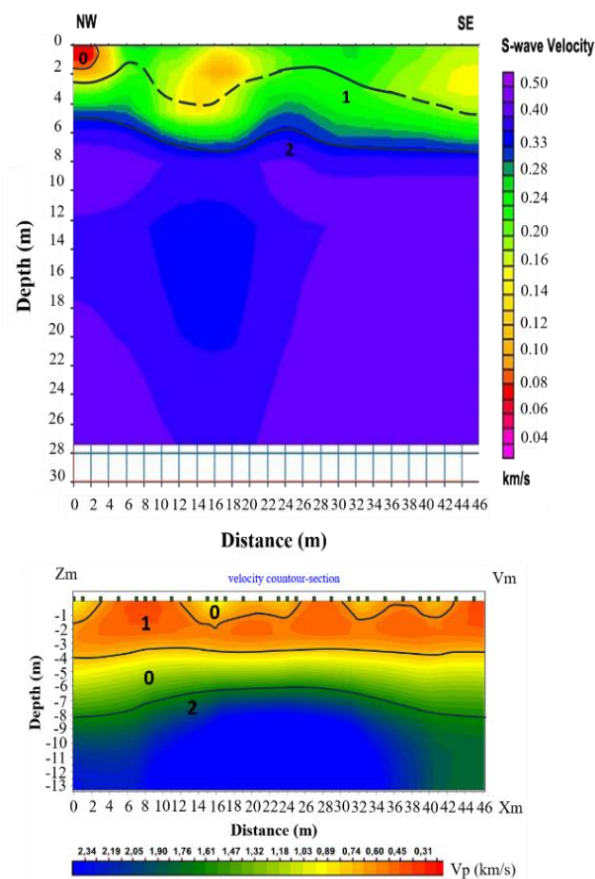


Figure 3. The 2D subsurface image from refraction seismic and MASW data

3.2 S-Wave Velocity

MASW seismic data processing is based on the concept of wave dispersion which through the inversion process in Seis-Imager software can be interpreted as 3 rock layers ranging in shear wave velocity (V_s) between 60 m/s to 400 m/s. MASW seismic subsurface images are also shown with a color bar scale that can be interpreted based on reference values such as Table 2. Based on that, it can be interpreted that the first layer is SE layers (soft soil) which has a velocity of less than 175 m/s which is named cluster 0 marked in red, the second layer is SD (medium soil) which ranges from 200 - 260 m/s which is named cluster 1 marked in green, and the third layer is SC (hard soil, very dense, soft rock) which ranges from 260 - 400 m/s which is named cluster 2 marked in purplish blue.

3.3 V_p/V_s Ratio

V_p/V_s Ratio means the ratio of P-wave velocity (V_p) to shear wave velocity (V_s) which can be used as an indicator in determining lithology. The s-wave velocity characteristic has a direction perpendicular to the direction of wave motion so that the combination with the V_p/V_s ratio can determine the difference in layers through which it passes. For example, soft rock has a relatively small s-wave velocity compared to hard rock because it is directly related to rock density [13]. The s-wave velocity (V_s) will weaken when passing through the fluid so that the soft rock may contain water content that affects the V_s value. By knowing the ratio of V_p and V_s , the water content in the rock layer can be estimated. In the research area from the surface to a depth of 12 meters the V_p/V_s ratio continues to increase by more than 2. It can be implied that rock hardness increases with depth.

3.3 Drilling Log Data

At the Krueng Kaleng Bridge location, a well has been drilled with drilling log data around 5 meters at a depth. Based on the report of the description of the borehole at a depth of 0 - 2 meters, a layer of gravelly coarse sand some clay is obtained, with a color of grayes yellow, with medium dense to dense, and has low to medium moisture content. At a depth of 2 - 5 meters, a layer of gravelly coarse sand mixed boulder some clay is obtained, grayes yellow color, with dense to very dense, and has medium to high moisture content. Refraction seismic and MASW interpretation data are also validated based on drilling log data and N-SPT values presented in Figure 4.

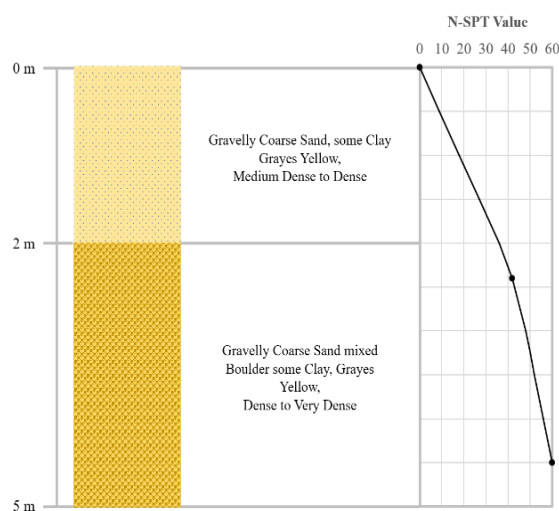


Figure 4. Drilling log data and N-SPT value

3.4 Clustering Analysis

Clustering analysis is done by connecting physical parameters with each other. As shown in Figure 5, each physical property that has been calculated has a close relationship with each other, represented by a number. The larger the number with a maximum number of 1.00 (100%), the closer the relationship between physical properties in the study area. In this analysis there is a very positive correlation between Vp/Vs ratio, Poisson ratio and density of 0.95, indicating a close

relationship between wave velocity and rock density.

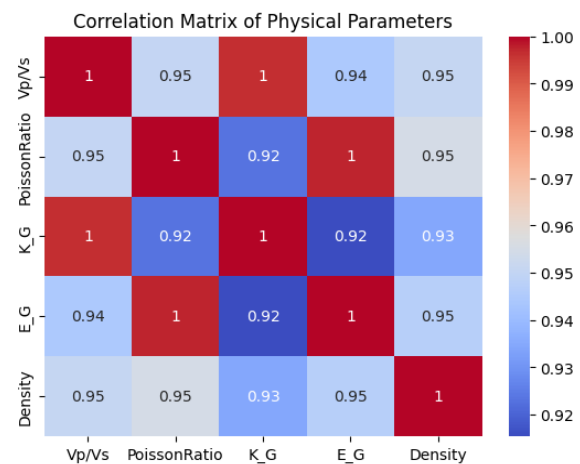


Figure 5. Percentage correlation matrix of interconnected physical

Poisson ratio itself is a property of rock stiffness or compactness that if a layer has pores in the rock, the poisson ratio value will decrease. In addition, there is a moderate relationship between Vp/Vs ratio and elastic modulus of 0.94, indicating that the elasticity of the rock is moderately affected by wave propagation. A prominent negative correlation is found between poisson ratio and bulk modulus (K), indicating that if a material is difficult to compress, its poisson ratio tends to be low. It can be interpreted that if there is a brittle layer that has a low poisson ratio value compared to other layers, it has a high K value, allowing the existence of a hard or brittle layer that is easily deformed.

PCA analysis of the physical properties data of the study area resulted in three clusters of lithological zones. PC1 is the main component that contributes a lot of essential information to the input data. PC1 explains 78% of the data variation, while PC2 explains 15% of the data variation, if the total is 93% which is quite representative of the original data as a whole. In Figure 6, it can be seen that there are points that are quite close to each other, which means they are similar in physical properties.

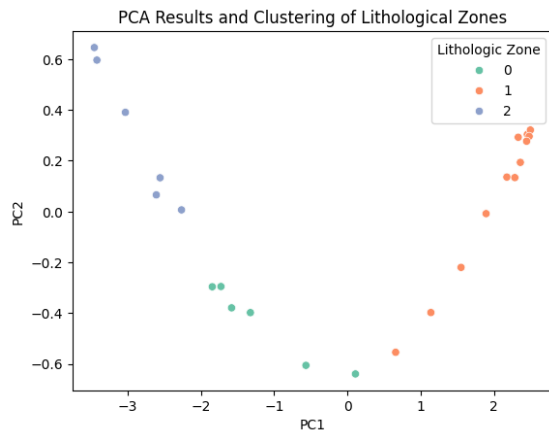


Figure 6. Principal Component Analysis (PCA) and Clusterin of Lithological

After analyzing PCA and K-Means and classifying the lithological zones, as shown in Figure 6 there are 3 types of lithological zones in the depth range of 0 to 13 meters. Cluster 0, which is soft green in color, is in two depth ranges, namely 0 to 1 meter, then the other ranges in depth from 5 to 6.3 meters. Cluster 1 is characterized by a gray color in the depth range of 1.8 - 4.2 meters. While Cluster 2 is characterized by a bright green color at a depth of > 6.3 meters.

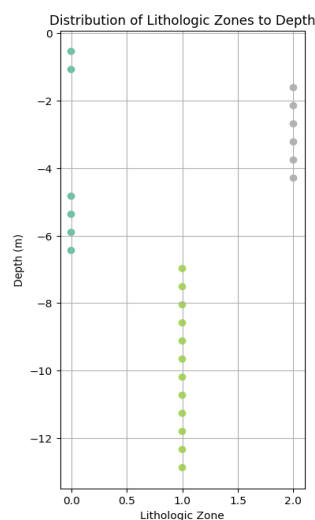


Figure 7. Distribution of lithologic zone after PCA and K-means process

Based on Figure, it can be seen that there is a positive relationship between the physical parameters and the depth and interpretation of the lithologic zone. As illustrated in the 2D subsurface cross section, there are three

clusters of lithologic zones where in the near-surface area there are hard layers that are not continuous over softer layers. The same can be seen in the shifting curves with decreasing parameter values as the near-surface physical parameters start with high values.

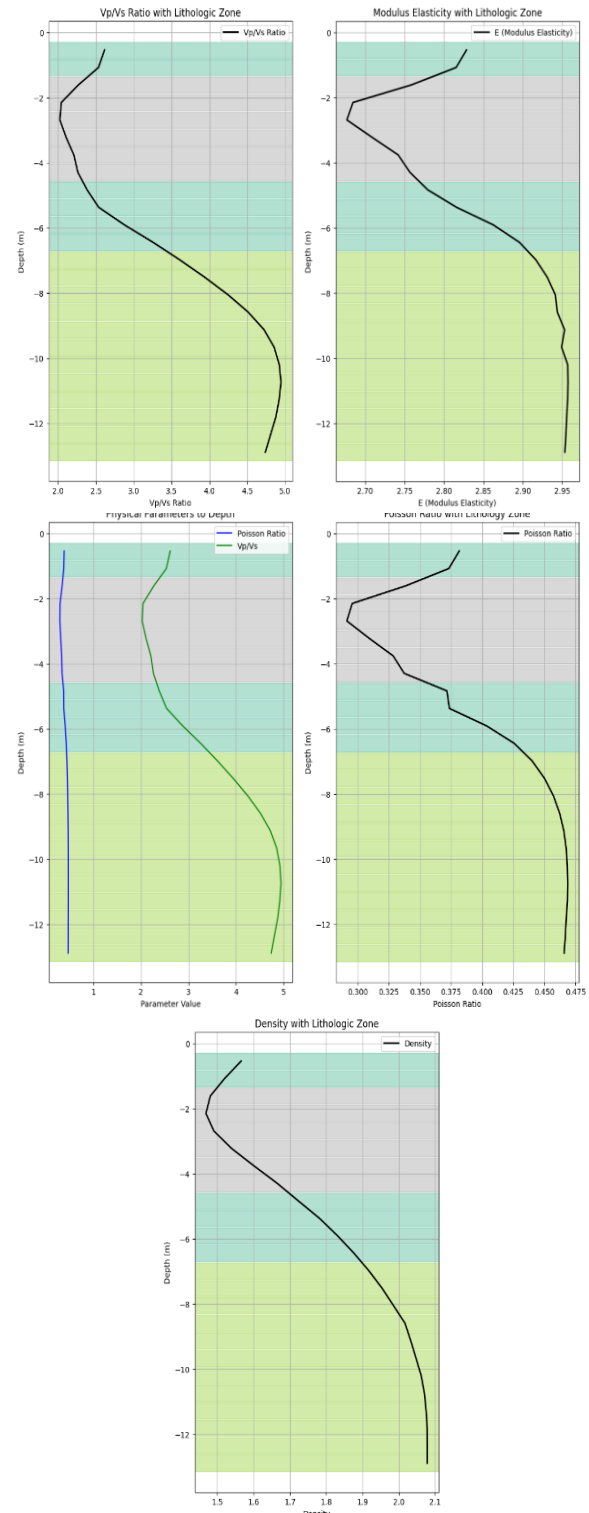


Figure 8. Relationship between physical parameters with dept and lithologic zones

Each parameter is presented with the same lithology type which has been explained previously that there are 3 clusters marked with soft green, gray, and bright green colors. First, a high V_p/V_s ratio generally indicates materials with higher fluid content or softer lithologies, such as clay-rich sediments or saturated layers. Thus, lower V_p/V_s ratio in the middle depth may correspond to more consolidated. The V_p/V_s ratio increases if V_s drops relatively more than V_p . This is in line with the log data where the deeper towards 5 meters, the more boulders that allow the density to be small so that they contain water so that V_s becomes small because it is unable to travel through the medium of cluster 2.

Then, low modulus near the surface indicates unconsolidated or weakly cemented sediments that deform easily under stress. Higher modulus in deeper layers corresponds to lithologies that are more compacted, denser, and resistant to deformation. High poisson's ration can indicate water-saturated or clay-rich materials, while lower values often correspond to dry, stiff, or brittle rock. So, water saturation and clay content increase the ratio, while more rigid lithologies like volcanic rocks lower it. Lastly, density increases consistently with depth, with the lowest densities at the surface and the highest in the deepest lithologic zone.

In shallow zone have low density, high poisson's ratio, and high V_p/V_s that suggesting unconsolidated sediments or weathered material, likely clay-rich and saturated. Then, in deep zone have a high density, high modulus that consistent with well-lithified, strong and brittle rocks.

3.5 Discussion

Previous research at the research area has been carried out using geoelectrical methods by [14], which at a depth of 0 – 5 meters is interpreted as clay sand. In addition, the study

obtained several layers of clayey sand (resistivity value $5\Omega m - 50\Omega m$), limestone, clay and the presence of weathering zones (resistivity value $5\Omega m - 200\Omega m$). Based on the interpretation of seismic velocities that actually use geophysical methods, it also shows the suspected saturated and unsaturated clayey sand layers.

Based on the results of V_p and V_s wave velocities obtained from geophysical methods and comparison of clustering results using PCA and K-means analysis through the Google Colab platform, three layers of rock can be distinguished, namely top soil, saturated clayey sand, and unsaturated clayey sand. If based on this interpretation, it can be concluded that the characteristics of the wave velocity are getting deeper and higher so that the layer is more compact, and if there is a shifting curve experienced, it is likely due to the existence of a layer that is still hard to survive from high weathering activity, characterized by the adjacent layer being top soil.

If it is associated with geological map data that the research area is included in the old Bentaro Volcanic Formation (Muvb) which is Jurassic to Cretaceous with basalts, agglomerate and intrusion lithologies, then now at a depth near the surface there is a clayey sand which indicates alteration and weathering so as to produce secondary lithologies such as clay. This can be validated by the borehole data which also contains sand and clay.

Due to ongoing geological processes to date, loose material may become new sediments that will be deposited in one river basin, namely Krueng Kaleng Bridge, to explain the current sedimentation process. Therefore, the recommendation for foundation construction based on the results of this layer interpretation is better.

4. CONCLUSION

- a. Based on the results of data processing, three layers of rock were obtained, namely top soil, saturated clayey sand, and unsaturated clayey sand. This is similar to the results of running machine learning by analyzing PCA and K-Means for clustering lithological zones, namely 3 zones. The interpretation results based on the reverence values are also similar to the borehole data so that the interpretation can be validated.
- b. The use of machine learning platforms with Python code commands for clustering analysis is considered effective for validating data and interpretation patterns, especially in the identification of lithological zones based on physical properties.
- c. According to the interpretation findings, the author suggests that a foundation for the construction of the Krueng Kaleng bridge should be placed at a depth greater than 20 meters since the upper layer consists of clayey sand, which was unable to support the foundation of the Jantho-Lamno bridge connection.

5. ACKNOWLEDGMENTS

The authors would like to thank CV. Dinar Geologi, Civil Engineering Division, for providing the drilling log data, as well as the supervisors and colleagues for their guidance, participation, and prayers.

6. REFERENCES

- [1] J. D. Bennet *et al.*, "Geologic map of the BANDA ACEH quadrangle, north sumatra, Scale 1:250,000," 1981, *Geological Research and Development Centre*.
- [2] D. A. Otchere, *Data Science and Machine Learning Applications in Subsurface Engineering*. 2024. doi: 10.1201/9781003366980.
- [3] M. Verleysen and M. Verleysen, "Principal Component Analysis (PCA)," *Statistics (Ber)*, no. September, pp. 1–8, 2001, doi: 10.5455/ijlr.20170415115235.
- [4] M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," *Nature Reviews Methods Primers*, vol. 2, no. 1, p. 100, 2022, doi: 10.1038/s43586-022-00184-w.
- [5] I. Isfan, A. Haris, A. Harsono, and A. Haris, "Cluster Analysis of Lithology Grouping Trends using Principal Component Spectral Analysis and Complex Seismic Attributes," *Makara J Sci*, vol. 25, no. 1, 2021, doi: 10.7454/mss.v25i1.1227.
- [6] H. Ugalde and W. A. Morris, "Cluster analysis of Euler deconvolution solutions: New filtering techniques and geologic strike determination," *Geophysics*, vol. 75, no. 3, pp. 61–70, 2010, doi: 10.1190/1.3429997.
- [7] M. C. de Matos, M. (Moe) Yenugu, K. J. Marfurt, and S. M. Angelo, "Channel delineation and chert reservoir characterization by integrated seismic texture segmentation and cluster analysis," pp. 801–806, 2011, doi: 10.1190/sbgf2011-165.
- [8] R. S. Kamat, "Earthquake Cluster Analysis: K - Means Approach," *Journal of Chemical and Pharmaceutical Sciences JCPS*, vol. 10, no. 1, pp. 250–253, 2017.
- [9] Z. Fadhli *et al.*, "Ground Surface Quality Assessment Using P-wave Velocity from 2-D Seismic Refraction Method," *Aceh International Journal of Science and*

- Technology*, vol. 11, no. 3, pp. 258–265, 2022, doi: 10.13170/aijst.11.3.28818.
- [10] Badan Standardisasi Nasional (BSN), *Metode uji CBR laboratorium (SNI 1744:2012)*. Jakarta: BSN, 2012.
- [11] J. P. Castagna, M. L. Batzle, and R. L. Eastwood, “Relationships Between Compressional-Wave and Shear-Wave Velocities in Clastic Silicate Rocks.,” *Geophysics*, vol. 50, no. 4, pp. 571–581, 1985, doi: 10.1190/1.1441933.
- [12] Handika, “Pemanfaatan Python dan Google Colab Dalam Pembelajaran Statistika Deskriptif,” *Edumatnesia: Prosiding Seminar Nasional Matematika dan Pendidikan Matematika*, pp. 379–389, 2024.
- [13] T. Yulianto, D. P. Sasongko, G. Yulianto, R. D. Indriana, A. Setyawan, and S. Widada, “Correlation of Vp/Vs ratio against the resistivity value to determine the aquifers presence estimation in jetak sub-village, getasan sub-district, Semarang regency,” *Journal of Physics: Conference Series*, vol. 1943, no. 1, 2021, doi: 10.1088/1742-6596/1943/1/012028.
- [14] R. Gemasih, “Identifikasi Struktur Bawah Permukaan Menggunakan Metode Geolistrik Resistivitas dan Induced Polarization (IP) Pada Area Pembangunan Jembatan Krueng Kaleng, Sabet, Aceh Jaya,” Universitas Syiah Kuala, 2016.