A Machine Learning Approach to Estimate Geo-mechanical Parameters from Core Samples: A Comparative Approach

JWNGSAR BRAHMA
Department of Mathematics, School of Technology,
Pandit Deendayal Energy University,
Rysan, Gandhinagar-382421,
INDIA

Abstract: - Geo-mechanical parameters and Thomsen parameters play very important roles to design a stable wellbore in a challenging environment. The main objective of this paper is to estimate Thomsen's parameters (ε, γ, δ) and geo-mechanical properties from the core samples by Machine Learning and a comparative analysis with the conventional mathematical approach; to place emphasis on the use of Machine Learning and Artificial Intelligence in the Oil & Gas industry and to highlight its future potential to help in the digital transformation of the industry. Two different Machine Learning models, the Ordinary Least Square method and the Random Forest method, were used to predict the aforementioned geo-mechanical properties from the wave velocity and confining pressure data. In this study, it has been observed that the approaches employed in the estimate of geo-mechanical properties are rapid and reliable (about 93.5 percent accuracy) and may be applied in geo-mechanical modeling for wellbore stability analysis for safe and cost-effective well plan and design on a large scale. The analysis in this work indicates that Young's modulus and Poisson's ratio are heavily influenced by the anisotropy parameters. Finally, a comparison is made with mathematical approaches. The machine learning and artificial intelligence approaches shown here are excellently matched with mathematical approaches.
The geo-mechanical parameters and Thomsen parameters and be computed with reasonable accuracy with the help of our proposed ML algorithms. Our proposed ML model can predict the geo-mechanical parameters and Thomsen parameters from the velocity profile directly without complex mathematical computation. The mathematical model would have required us to first determine the stiffness constants for the prediction of that parameters.

Additionally, we may conclude that a machine learning model needs to be trained with more modeling data to predict the right values with a smaller error margin. The number of data points required to train a model has a significant impact on the model's overall accuracy. Therefore, additional modeling data is needed to learn about and comprehend the intricacies, patterns, and interactions between provided input and output variables.

Key-Words: - Machine Learning, Artificial Intelligence, Core Samples, Thomsen Parameter, Geo-mechanical.

Received: February 2, 2023. Revised: May 25, 2023. Accepted: June 17, 2023. Published: July 18, 2023.

1 Introduction

The anisotropy in a rock is defined as its properties that vary with the direction of observation. This anisotropy may present everywhere within the subsurface. It is the different geological origins of each rock that creates different sedimentary features in them, [1]. These properties are often measured parallelly or perpendicularly to the sedimentary features (bedding planes), with the earth's anisotropic response to changes investigated with the help of seismic, sonic, or ultrasonic surveys, [2].

Fundamental geo-mechanical properties include stress, strain, Young's modulus, Poisson's ratio, and compressive strength. Geo-mechanical evaluation is required in petroleum engineering for rock failure prediction, determination of in-situ stress, wellbore stability analysis, hydraulic fracturing design, and anisotropy measurement. Geo-mechanical rock properties are a subset of petrophysical parameters that can be calculated directly in rock mechanics labs or field experiments, [3]. However, since they are more or less highly correlated with other petrophysical parameters (e.g., elastic wave velocities), an "indirect" derivation from geophysical measurements is being researched and applied.

Three parameters characterize anisotropy, in addition to the normal V_P, V_S, and ρ. Thomsen's parameters include δ (delta), ϵ (epsilon), and γ (gamma), [4]. The short offset effect, δ or delta, captures the relationship between the velocity

required to flatten gathers (the NMO velocity) and the zero-offset average velocity recorded by check shots. It is simple to calculate, but it may be challenging to comprehend physically. ϵ or epsilon, the long offset impact is "the fractional difference between vertical and horizontal P velocities, i.e., it is the parameter commonly referred to as 'the' anisotropy of a rock." Horizontal velocity, on the other hand, is difficult to measure, [4]. The shear wave effect, often known as γ or gamma, contrasts a horizontal shear wave with horizontal polarisation to a vertical shear wave.

Shales comprise about 70% of sedimentary basins. However, due to the friable nature of shales, there are very few laboratory measurements of velocity anisotropy, [5], [6], [7], [8]. More recently, the influence of pore fluid on the elastic properties of shale has been investigate by Hornby in 1998. He measured compressional and shear wave velocities up to 80MPa on two fluid-saturated shale samples under drained conditions, [9]. One sample was Jurassic outcrop shale that was recovered from undersea and stored in its natural fluid, and the other is Kimmeridge clay taken from a North Sea borehole. Measurements were made on cores parallel, perpendicular, and at 45^0 to bedding. Values of anisotropy were up to 26% for compressional and 48% for shear wave velocity and were found to decrease with increasing pressure, [10]. The effect of reduced porosity was therefore concluded to be more influential on anisotropy than an increased alignment of minerals at higher pressure. The elastic constants, velocities, and anisotropies in shales can be obtained from traditionally measured on multiple adjacent core plugs with different orientations, [10]. To derive the five independent constants for transversely isotropic (TI) rock, Wang measured three plugs separately (one parallel, one perpendicular, and one $\pm 45^0$ to the symmetric axis). The advantage of this three-plug method is redundancy for the calculation of the five independent elastic constants since each core plug measurement yields three velocities. Unlike shale, clean sandstone is intrinsically isotropic. Sandstones are rarely clean; they often contain minerals other than quartz, such as clay minerals which can affect their reservoir qualities as well as their elastic properties. The presence of clay minerals and clastic sheet silicates strongly influences the physical and chemical properties of both sandstones and shales, [11]. Clay can be located between the grain contacts as structural clay in the pore space as dispersed clay or as laminations, [12]. The distribution of the clay will depend on the conditions at deposition on compaction, bioturbation, and diagenesis. While most reservoirs are composed of relatively isotropic sandstones or carbonates, their properties may be modified by stress. Non-uniform compressive stress will have a major effect on randomly distributed microcracks in a reservoir. When the rock is unstressed all of the cracks may be open, however, compressional stresses will close cracks oriented perpendicular to the direction of maximum compressive stress, while cracks parallel to the stress direction will remain open. Elastic waves passing through the stressed rock will travel faster across the closed cracks (parallel to maximum stress) than across the open ones.

The effects of anisotropy in seismic data will reveal a lot about the Earth's processes and mineralogy. Seismic anisotropy has garnered a lot of attention from academia and business in the recent two decades. Many seismic processing and inversion procedures now use anisotropic models, which provide significant improvements in seismic imaging efficiency and resolution. The employment of an anisotropic velocity model in conjunction with seismic imaging has greatly reduced the uncertainty surrounding internal and bounding-fault locations, reducing the likelihood of making an investment choice simply based on seismic interpretation.

In addition, the discovery of a link between anisotropy parameters, fracture orientation, and density has resulted in the development of realistic reservoir characterization methodologies. If fractures are considered throughout the drilling choice process, the drainage area of each producing well can be greatly expanded, due to the acquisition of such parameters as fracture spatial distribution and density. The drilling cost of exploration and production (E&P) projects will be greatly decreased because there will be fewer wells due to the expanded drainage area.

One of the most critical aspects of preparing a strategy for hydraulic fracking is understanding the geo-mechanical rock properties. Reduce operational risk and optimize production while spending as little money as possible, especially in ultra-tight complicated formations like shale, where operational risk is significant owing to formation uncertainty. As a result, machine learning and artificial intelligence (AI) are becoming important in the oil business, as they generate accurate information by merging log and core data. This method is essential for forecasting the geo-mechanical parameters of shale, the most heterogeneous rock with the least desired wettability for hydrocarbon flow. As a result,

machine learning and artificial intelligence (AI) can aggregate and correlate data more accurately than a human can. Manual integration and correlation are less reliable, effective, and, most crucially, less expensive and time-consuming. Many researchers have focused on combining core and log data to quantify geo-mechanical parameters and output sweet spots in the Eagle Ford and Barnett formations using machine learning and artificial intelligence, [13].

Wave velocities, Poisson's ratio, Young's, shear, and bulk modulus are all significant rock mechanical properties in the geo-mechanical study of petroleum reservoirs. Due to the unavailability of data or the steep costs for testing, direct measurement of these parameters is typically impossible, particularly in older wells. Hence, to predict these parameters from available data, indirect methods are frequently used. Empirical equations are the most basic and widely used tool. These correlations, on the other hand, are very susceptible to various types of fluids or lithology and are frequently unrelated to local geology. In recent years, intelligent systems have been applied in a variety of fields of science and technology, and they have consistently proven to be useful in prediction and optimization challenges, [14].

Challenges that the industry currently faces include complex mathematical modeling that is extremely difficult and time-consuming. Apart from that, several complex calculations are performed to calculate the stiffness constant using the Voigt matrix, which can have dimensions up to 81 x 81. The time needed to solve the matrix of such large dimensions is extraordinarily complex and takes several months for the solution to be reached. These challenges can be easily overcome using AI/ML to calculate geo-mechanical and anisotropic parameters, giving reasonably high accuracy.

In this research work, data from lab studies of four types of cores – Dry Sandstone, Shale, Sandy Shale, and Saturated Sandstone – was utilized. The goal of this study is to determine how elastic anisotropy affects Young's modulus and Poisson's ratio. V_P, V_{Sh}, and V_{SV} wave velocity data were measured as a function of confining pressure from 1 MPa to 40.3 MPa, obtained through the ultrasonic transmission method. The study aims to predict here stress, strain, Young's modulus, Poisson's ratio, and Thomsen's parameters using a data-driven approach with the help of Machine Learning algorithms. Pandas (W. McKinney and Pandas) and Numpy (T. Oliphant) library of the phyton are used in this computation. During the prediction of these Geomechanical parameters we used V_P, V_{Sh}, and V_{SV} wave velocity and confining pressures as an input and we predicted Thomsen's parameters, Young's modulus, and Poisson's ratio as an output.

2 Methodology

2.1 Mathematical Analysis

Core data (velocity and density) of four different types of sedimentary rock is taken to calculate Thomsen's anisotropic parameters and geo-mechanical properties concerning different confining pressure. Cores of different rock types were used for the study of dry sandstone, shale, sandy shale, and saturated sandstone from a particular basin.

Transverse isotropy is commonly seen in sedimentary rocks. Each layer has similar qualities in-plane but distinct properties as it progresses in thickness. Each layer's plane is the isotropy plane, and the vertical axis is the symmetry axis. The horizontal velocity differs from the vertical velocity in vertical transverse isotropy, [15].

For the measurement of transverse isotropy, the standard three-plug method is used. Using the standard three-plug method, each rock sample is cut in three different orientations: parallel, perpendicular, and typically 45˚ to the vertical symmetry axis, [15]. The details of measurement methods are shown in Figure 1 and Figure 2 (Appendix). Phase velocity measurement for each core sample by two orthogonal and one compressional shear wave is performed using dynamic or static ultrasonic laboratory measurements. So, each core sample's total three velocities are determined; therefore, each rock sample's total of nine velocities is determined.

In vertical transverse isotropy (VTI), the following equation describes the relationship between strain (e_{kl}) and stress (σ_{ij}):

$$\sigma_{ij} = C_{ijkl} e_{kl} \tag{1}$$

where C_{ijkl} represents the Voigt matrix, [16]. If the vertical axis is noted by Z, then the other two principal axes (X and Y) are parallel to the transversely isotropic plane.

In linear elasticity, Hooke's law states that the stress and strain are related, i.e.

$$\sigma = Ce \tag{2}$$

Here stress is in N/mm^2 and strain and elastic constants are dimensionless quantities.
Here σ is stress, e is strain and C is elastic constant and are defined as below:

$$\sigma = \begin{bmatrix}\sigma_1\\ \sigma_2\\ \sigma_3\\ \sigma_4\\ \sigma_5\\ \sigma_6\end{bmatrix}; e = \begin{bmatrix}e_1\\ e_2\\ e_3\\ e_4\\ e_5\\ e_6\end{bmatrix} \text{ and}$$

$$C = \begin{bmatrix}C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16}\\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26}\\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36}\\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46}\\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56}\\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66}\end{bmatrix}$$

The equation (2) becomes as below:

$$\begin{bmatrix}\sigma_1\\ \sigma_2\\ \sigma_3\\ \sigma_4\\ \sigma_5\\ \sigma_6\end{bmatrix} = \begin{bmatrix}C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16}\\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26}\\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36}\\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46}\\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56}\\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66}\end{bmatrix}\begin{bmatrix}e_1\\ e_2\\ e_3\\ e_4\\ e_5\\ e_6\end{bmatrix} \quad (3)$$

For Vertical isotropic material, this Voigt matrix is reduced to only five non-zero independent elastic constants, which are C_{11}, C_{33}, C_{44}, C_{66}, and C_{13}.

$$C = \begin{bmatrix}C_{11} & C_{11}-2C_{66} & C_{13} & 0 & 0 & 0\\ C_{11}-2C_{66} & C_{11} & C_{13} & 0 & 0 & 0\\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0\\ 0 & 0 & 0 & C_{44} & 0 & 0\\ 0 & 0 & 0 & 0 & C_{55} & 0\\ 0 & 0 & 0 & 0 & 0 & C_{66}\end{bmatrix} \quad (4)$$

From the equation (3), Voigt matrix C has five non-zero independent elastic constants: C_{11}, C_{33}, C_{44}, C_{66}, and C_{13}. The sixth elastic constant is $C_{12} = C_{11} - 2C_{66}$, where,

C_{11} = in-plane compressional modulus,
C_{33} = out-of-plane compressional modulus,
C_{44} = out-of-plane shear modulus,
C_{66} = the in-plane shear modulus,
C_{13} = important constant that controls the shape of the wave surfaces.

It is essential to establish a relationship between five non-zero independent elastic constants with geo-mechanical and Thomsen parameters. We need to use a compliance matrix (inverse of the elastic stiffness matrix) to establish this relationship. The elastic constant can be estimated from phase velocity data.

For a hexagonal material (VTI material), two dynamic Poisson's ratios can be obtained using five elastic stiffness C_{ij}, one vertical Young's modulus (E_3), and one horizontal Young's modulus (E_1), [17]:

$$E_1 = \frac{[C_{33}(C_{11}+C_{12})-2C_{13}^2](C_{11}-C_{12})}{C_{33}C_{11}-C_{13}} \quad (3)$$

$$E_3 = \frac{[C_{33}(C_{11}+C_{12})-2C_{13}^2]}{C_{11}+C_{12}} \quad (4)$$

$$\vartheta_{31} = \frac{C_{13}}{C_{11}+C_{12}} \quad (5)$$

$$\vartheta_{12} = \frac{C_{33}C_{12}-2C_{13}^2}{C_{11}C_{33}-2C_{13}^2} \quad (6)$$

These dynamic Poison's ratios ϑ_{ij}, are indirect measures of the ratio of the lateral to axial strains when the uniaxial stress is applied in the same direction of axial strain.

Using five elastic stiffness C_{ij}, Thomsen's parameters can be estimated by the following mathematical equations suggested by Thomsen (1986):

$$\epsilon = \frac{C_{11}-C_{33}}{2C_{33}} \quad (7)$$

$$\delta = \frac{(C_{13}+C_{44})^2-(C_{33}-C_{44})^2}{2C_{33}(C_{33}-C_{44})} \quad (8)$$

$$\gamma = \frac{C_{66}-C_{44}}{2C_{44}} \quad (9)$$

2 - 4

2.2 Machine Learning Approaches

2.2.1 Data Preparation

For machine learning applications, once the measurements data from the core samples are collected there is a complete procedure that must be carried out before training the machine learning model.

1. **Data Selection**

The purpose of this stage is to choose a subset of all available data. In this step, *pandas* and *the numpy* library of Python are used. In this process, the most considerable data has been selected which is best suitable for our problem statement.

2. Data Pre-Processing

Three common data pre-processing steps include:

- **Formatting:** It's possible that the data that is chosen isn't in the right format to work with. As a result, the data has been converted to an Excel spreadsheet.
- **Cleaning:** The removal or correction of missing data is known as data cleaning. Some data instances may be incomplete, and they'll need to be eliminated. Additionally, some of the attributes may contain sensitive information, and these attributes may need to be anonymized or eliminated from the data. The outlier removal method has been used to eliminate outliers from the data using the is null () function for the null value of the data set.
- **Sampling:** There may be a lot more selected data than is required. More data can lead to substantially longer algorithm execution times as well as increased computing and memory needs. As a result, before evaluating the entire dataset, a smaller subset sample of the selected data works considerably faster for exploring and developing ideas.

3. Data Transformation

Scaling, attribute decompositions, and attribute aggregations are three popular data transformations. This process is called feature engineering.

- **Scaling:** The pre-processed data could have a mix of scales for different quantities like dollars, kilograms, and sales volume. Many machine learning algorithms prefer data attributes with the same scale, such as 0 to 1 for the smallest and highest value for a specific feature. Consider any feature scaling that may be required.
- **Decomposition:** There may be features that indicate a complicated notion that, when broken down into basic bits, are more valuable to a machine learning method. A date, for example, may comprise day and time components that could be separated further. Perhaps only the time of day has any bearing on the problem at hand.
- **Aggregation:** There may be features that can be combined into a single feature to make the problem more meaningful. For example, each time a client logged into a system, there may be data instances for the number of logins, allowing the extra instances to be discarded. Consider the several types of feature aggregations that could be used.

4. Outlier Detection and Removal

Outliers are extraordinary results that differ significantly from the rest of the data. Outliers in a normal distribution. Outlier mining, outlier modeling, novelty identification, and anomaly detection are all terms used in data mining and machine learning to describe the process of discovering outliers.

Here are some outlier detection methods:

- **Extreme Value Analysis:** Calculate the statistical tails of the data's underlying distribution univariate data.
- **Probabilistic and Statistical Models:** Determine unlikely events using a probabilistic data model. Gaussian mixture models were optimized via expectation maximization.
- **Linear Models:** Methods for projecting data into lower dimensions based on linear correlations. Principal component analysis (PCA) and data with substantial residual errors may be considered outliers.
- **Proximity-based Models:** Cluster, density, or closest neighbor analysis is used to isolate data instances from the rest of the data.
- **Information Theoretic Models:** Outliers are data occurrences that add to the complexity of a dataset (minimum code length).
- **High-Dimensional Outlier Detection:** Distance-based metrics in higher dimensions are broken down using methods that explore subspaces for outliers. (Curse of dimensionality).

2.2.2 Data Splitting

Taking a dataset and separating it into two subgroups is the technique. The training dataset is the first subset, which is used to fit the model. The second subset is not used to train the model; instead, the dataset's input element is given to the model, which then makes predictions and compares them to the predicted values. The test dataset is the name given to the second dataset.

- **Train Dataset:** Used to train by feeding to the machine learning model.

- **Test Dataset:** Used to evaluate the metrics of an already trained machine learning model and to check for accuracy.

The size of the train and test sets, which is usually stated as a percentage between 0 and 1 for either the train or test datasets, is the procedure's key configuration parameter. Here in the training set, we consider a size of 80% of the total samples 1000 and test set with the remaining 20%.
Here, using the *train_test_split()* syntax of Python, the data has been split into the train data set and test dataset.

2.2.3 Dummy Data

The amount of available original data is not sufficient to train the machine learning model, so instead, we created dummy data (1000 samples) with the help of Python functions using the original data. The Python function has some basic input arguments which use the minimum, maximum, mode value, and the quantity of the data points to be generated. Suppose we want 50 dummy data points of the available original data; we need to input the fundamental above-listed argument values. The function generates values that can be utilized better to understand the complexities of our machine learning model.

2.2.4 Data Visualization

In applied statistics and machine learning, data visualization is a must-have talent. It can be useful for identifying patterns, faulty data, outliers, and other things when studying and getting to know a dataset. Data visualizations can express and demonstrate crucial links in plots and charts that are more visceral to yourself and stakeholders than measurements of association or importance with just a little subject knowledge. The *matplotlib()* and *seaborn()* libraries of Python are used for the data visualization.

2.2.5 Machine Learning Model Selection

Once the Excel file with all the features (confining pressure and shear waves in different directions) variables and output (Thomsen's parameters) variables have been imported, then comes the data pre-processing, which is a very crucial step that helps us understand the data and any flaws that are associated with it which may affect the overall accuracy of the machine learning model.

To make predictions, visualization is an important step that gives us a graphic representation of our data. When we have very numerous data, by just looking at the numbers, we cannot interpret anything out of it unless we visualize the data in a graphical representation. For instance, if we have some density-neutron log data, just looking at numbers would not help the geologist drive important decisions. When we put these data into a graphical visual context, we can have a better understanding of the logs and their features (for example, the crossovers). So, visualization helps to know the trends of the data, the patterns, and the outliers.

Accuracy is a crucial parameter for the selection of the machine learning model. However, our focus of the study is to predict the Thompson parameter values with given shear wave values in a different direction, and it falls under the linear regression problem. In the regression problem/model, there are specific evaluation metrics. Let's have a look at some critical evaluation metrics.

- **R-squared value (R^2):** R^2 is a statistical measure of the level of the correlation between the observed outcome and the predicted value given by the model. So, if a model achieves an R^2 score of 1, then it can be understood that both variables are perfectly correlated to each other, which implies no variance. In another way, (total variance explained by the model)/ (total variance) the value of this equation signifies the quality of the correlation between the variables. The closer the value of R^2 to 1, the better the model is considered to fit.
- **Adjusted R-squared:** It's a modified version of R-squared that adjusts for predictors in a regression model that are no longer significant. It demonstrates whether or not increasing the number of independent variables improves the model. The Adjusted R-squared value is always less than the R-squared value.
- **Root Mean Squared Error (RMSE):** Another popular method is where the regression prediction errors are calculated. It's essentially the average of the squared errors or the difference between the dataset's observed value and the model's predicted value. The square root of the mean squared error (MSE), which is the average squared difference between the observed actual outcome values and the values predicted by the model, is the RMSE in mathematics. The better the fit, the lower the RMSE number.
- **Mean Absolute error (MAE):** It's similar to RMSE, except that MAE gauges the model's prediction error. It is the average absolute difference between observed and projected results in mathematics. MAE is somewhat unaffected by outliers in the sample.

R-squared tells us how the independent variables explain much variance in the dependent variable. In a general way, if we add more amount of observations and more independent variables, the value of the R-squared increases, but when the R-squared value does not increase any further with adding more independent variables, we need to understand that the added variables are uncorrelated with the dependent variable. In most complex predictions, it is recommended to use Adjusted R-squared in place of R-squared for model evaluation as it gives some penalty for an extra variable if the previous variable does not explain the dependent variable more correctly. However, in our case, we have a less complex situation, so it is better to use the R-squared method for model evaluation and selection in our study.

2.2.6 Algorithms Used in Machine Learning

- **Ordinary Least Squares (OLS)**

For estimating the unknown parameters in a linear regression model, ordinary least squares (OLS) is a sort of linear least-squares method. By minimizing the sum of the squares of the differences between the observed dependent variable (values of the variable being observed) in the given dataset and those predicted by the linear function of the independent variable, OLS chooses the parameters of a linear function of a set of explanatory variables.

Geometrically, the total of the squared distances between each data point in the set and the corresponding point on the regression surface, measured parallel to the axis of the dependent variable—the lower the differences, the better the model fits the data. The resulting estimator, especially in a basic linear regression with a single regressor on the right side of the regression equation, can be stated by a simple formula.

The OLS coefficient estimates for the simple linear regression are as follows:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_0 \bar{x} \tag{10}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{11}$$

where the “hats” above the coefficients indicate that it concerns the coefficient estimates, and the “bars” above the x and y variables mean that they are the sample averages, which are computed as follows:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{12}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{13}$$

- **Random Forest**

The bagging approach is used to train Random Forests. Bagging, also known as Bootstrap Aggregating, entails picking subsets of the training data at random, fitting a model to these smaller data sets, and then aggregating the results. Given that we are sampling with replacement, this strategy permits numerous instances to be utilized again for the training step. Sampling subsets of the training set, fitting a Decision Tree to each, and aggregating the results is what tree bagging is all about. The systematic diagrammatic representation of the Random Forest is shown in Figure 4 (Appendix). By applying the bagging approach to the feature space, the Random Forest method offers more randomness and diversity. Instead of searching greedily for the best predictors to construct branches, it randomly samples elements of the predictor space, increasing variety and lowering variance while maintaining or increasing bias. This strategy is also known as "feature bagging," and it leads to a more robust model.

Random Forests is an algorithm where each new data point goes through the same process, but in the ensemble, it visits all the different trees, which were grown using random samples of both training data and features. The functions for aggregation will differ according to the task at hand. It uses the mode or most frequent class predicted by the individual trees (also known as a majority vote) for classified problems, but in Regression tasks, the average prediction of each tree is used.

Building and Training Random Forest Models with Scikit-Learn

Assuming any two child nodes, Scikit – Learn calculates the importance of a node using Gini Importance for each decision tree.

$$N_{i_j} = W_j C_j - W_{left(j)} C_{left(j)} - W_{right(j)} C_{right(j)} \tag{16}$$

Where

$N_{i_j} = The\ importance\ of\ node\ j$

$W_j = Weighted\ number\ of\ samples\ reaching\ node\ j$

$C_j = The\ impurity\ value\ of\ node\ j$

$left(j) = Child\ node\ from\ left\ split\ on\ node\ j$

$right(j) =$
$Child\ node\ from\ right\ split\ on\ node\ j$

To calculate the importance of each feature on a decision tree, the following equation can be used.

$$f_{i_i} = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} N_{i_j}}{\sum_{k \in all\ nodes} N_{i_k}} \quad (17)$$

At the random forest level, the final feature importance by the following equation:

$$RFf_{i_i} = \frac{\sum_{j \in all\ trees}\ normf_{i_{ij}}}{T} \quad (18)$$

Where,
RFf_{i_i}
$= The\ importance\ of\ the\ feature\ i\ calculated\ from\ all\ trees\ in\ the\ Random\ forest\ model$
$normf_{i_{ij}}$
$= Normalised\ feature\ importance\ for\ i\ in\ tree\ j$
$T = total\ number\ of\ trees$
Where,
$f_{i_i} = The\ importance\ of\ feature\ i$
$N_{i_j} = The\ importance\ of\ node\ j$

It can be normalized to a value between 0 and 1 by the following formula:

$$normf_{i_i} = \frac{f_{i_i}}{\sum_{j \in all\ features} f_{i_j}} \quad (19)$$

Building and Training Random Forest Models with Spark

Spark determines the relevance of a feature by adding the gain multiplied by the number of samples that pass through the node for each decision tree as below:

$$f_{i_i} = \sum_{j:nodes\ j\ splits\ on\ feature\ i} S_j C_j \quad (20)$$

Where
$f_{i_i} = The\ importance\ of\ feature\ i$
$S_j = Number\ of\ samples\ reaching\ node\ j$
$C_j = The\ impurity\ value\ of\ node\ j$

The normalization for each tree can be estimated by the following equation:

$$normf_{i_i} = \frac{f_{i_i}}{\sum_{j \in all\ features}\ f_{i_j}} \quad (21)$$

Where
$normf_{i_i}$
$= The\ normalized\ importance\ of\ feature\ i$
$f_{i_i} = The\ importance\ of\ feature\ i$
Finally, the feature importance values for each tree can be calculated using the following equation:

$$RFf_{i_i} = \frac{\sum_j normf_{i_{ij}}}{\sum_{j \in all\ features, k \in all\ trees}\ normf_{i_{jk}}} \quad (22)$$

Where
RFf_{i_i}
$= The\ importance\ of\ the\ feature\ i\ calculated\ from\ all\ trees\ in\ the\ random\ forest\ model.$
$normf_{i_{ij}}$
$= The\ normalized\ features\ importance\ for\ i\ in\ tree\ j$
The accuracy of all the models is going to estimate with the parameter R^2. The graphical representation of R^2 and how it can be calculated is given in detail in Figure 3 (Appendix).

3 Results and Discussion

Dry sandstone, saturated sandstone, Shale, and Sandy shale are four types of rock samples that accurately describe elastic and anisotropic behavior in the sedimentary column. Because each of these rock types has a varied sensitivity to confining pressure, the anisotropies are influenced differently as the confining pressure changes.

The anisotropy parameters are plotted. Each meaning of the Thomsen anisotropy parameters is given as follows: The first parameter is γ; from the original equation, it is evident that γ only depends on the S-wave velocity component, whether fast or slow S-wave velocity. Hence, the parameter γ describes the anisotropy condition of the S-wave velocity. The anisotropy of the S-wave grows as the value of the anisotropy parameter γ increases. The second parameter is ε; ε describes P-wave velocity. The anisotropy of the S-wave grows as the value of the anisotropy parameter ε increases. Moreover, the third parameter is δ; δ describes both P-wave and S-wave. An increase in the anisotropy parameter δ indicates an increase in both P-wave and S-wave anisotropy.

Poisson's ratio (horizontal and vertical) and Young's modulus (horizontal and vertical) are important rock mechanical parameters in the geo-mechanical study of petroleum reservoirs. Due to the high expense of testing or a lack of relevant data, direct measurement of these parameters is frequently not possible, especially in older wells. As a result, machine learning models are frequently employed to forecast these parameters based on available data. Empirical equations are the most basic and widely used strategy. These correlations, on the other hand, are particularly sensitive to different types of fluids or lithology and are frequently irrelevant to local geology. In recent years, AI and Machine learning has been used in

different sciences and technologies and have often been demonstrated to help predict and optimize problems. Herein, a set of geo-mechanical parameters for different types of rocks has been predicted using Machine learning models (Ordinary Least Square method and Random Forest method). For this purpose, the mechanical properties of rocks, belonging to different lithologies, were predicted from wave velocities measured in the experimental studies on the core. The results depicted that the used methodologies were swift and reliable (93.5% accuracy) in the estimation of geo-mechanical properties and can be used in the geo-mechanical modeling of petroleum reservoirs on the industrial scale.

3.1 The Effect of Confining Pressure on Thomsen Anisotropic Parameters

- **Dry Sandstone**

With increasing confining pressure, the Thomsen anisotropic parameters show different trends. The P-wave anisotropy parameter ε increases linearly with increasing confining pressure and the S-wave anisotropy parameter γ, which shows a similar trend and increases in confining pressure. Furthermore, the third parameter is δ which describes that both P-wave and S-wave remain constant at zero and do not change with an increase in confining pressure (Figure 5, Appendix).

- **Sandy Shale**

With increasing values of confining pressure, the P-wave anisotropy parameter ε increases gradually, showing high values, while the S-wave anisotropy parameter γ also shows a similar trend, but with negative values, as it approaches zero. The third anisotropy parameter δ, which describes both P-wave and S-wave, remains constant at zero and does not change (Figure 6, Appendix).

- **Shale**

With increasing values of confining pressure, the P-wave anisotropy parameter ε increases gradually, showing high values, while the S-wave anisotropy parameter γ also shows a similar trend but with slightly lower values. The third anisotropy parameter δ, which describes both P-wave and S-wave, remains constant at zero and does not change (Figure 7, Appendix).

- **Saturated Sandstone**

With increasing values of confining pressure, the P-wave anisotropy parameter ε shows a relatively sharper decline before becoming more gradual. In contrast, the S-wave anisotropy parameter γ moves in the opposite direction, remaining constant initially before showing an upward trend. The third anisotropy parameter δ, which describes both P-wave and S-wave, remains constant at zero and does not change (Figure 8, Appendix).

3.2 Estimation of Geo-mechanical Parameters using Predictive Modelling Techniques and Comparison with Mathematical Model

Two different Machine Learning models, Ordinary Least Square (OLS) and Random Forest (RF), were used to predict horizontal & vertical Young's Modulus (E_1& E_3) and horizontal & vertical Poisson's Ratio (V_{12}& V_{31}) for the below samples. The predicted values were then compared with the values from the mathematical model, and calculated using empirical equations, to determine the accuracy.

- **Dry Sandstone**

While calculating E_1 and E_3 for a given set of confining pressures through the OLS method and Random Forest method, the OLS method gives an accuracy of approximately 98% and 94%, respectively, compensating the number of predictions only three out of six input values. Comparing it to the Random Forest method gives an accuracy of 92% and 98%, respectively, to predict the outcome for all six input values (Table 1, Appendix).

While calculating V_{12} and V_{31} for a given set of confining pressures through the OLS method and Random Forest method, the OLS method gives an accuracy of approximately 98% and 75%, respectively, compensating the number of predictions only three out of six input values. Comparing it to the Random Forest method gives an accuracy of 89% and 74%, respectively, to predict the outcome for all six input values in the first case (Table 2, Appendix).

For this particular sample, the RF method was vastly superior to the OLS method, and the obtained values were pretty accurate compared to those of the mathematical model.

- **Shale**

While calculating E_1 and E_3 for a given set of confining pressures through the OLS method, approximately 99% was obtained for both cases. The predicted values were found to be almost identical to those calculated using the mathematical model. RF method was not considered for the given Shale sample because the OLS method was deemed accurate enough (Table 3, Appendix).

While calculating V_{12} and V_{31} for a set of confining pressures through the OLS method, it gives an accuracy of approximately 98% and 97%,

respectively, which is almost identical to those calculated using the mathematical model (Table 4, Appendix).

RF method was not considered for the given Shale sample because the OLS method was deemed accurate enough.

- **Sandy Shale**

While calculating E_1 and E_3 for a given set of confining pressures through the OLS method and Random Forest method, the OLS method gives an accuracy of 98% for both cases, but with the slight drawback of accurately predicting outcomes, only three of the five input values. This irregularity is due to the complexity of calculating the constant elastic $C_{13,}$ which becomes a complex number at the associated pressures. On the other hand, the RF method faces no such complexities and can predict 97% and 99% accuracy for E_1 and $E_{3,}$ respectively (Table 5, Appendix).

While calculating V_{12} and V_{31} for a given set of confining pressures, the OLS method gives an accuracy of 68% and 78%, respectively, while the Random Forest method gives an accuracy of 99% and 74%, respectively. In V_{31} however, both the ML models were able to predict outcomes for only two of the five input values. This complication is due to the issue of data redundancy (Table 6, Appendix).

For this particular sample, the RF method was vastly superior to the OLS method, and the obtained values were reasonably accurate compared to those of the mathematical model.

- **Saturated Sandstone**

While calculating E_1 and E_3 for a given set of confining pressures, the OLS method accuracy is approximately 68% and 66%, respectively. Comparing it to the Random Forest method gives an accuracy of 92% and 94%, respectively (Table 7, Appendix).

While calculating V_{12} and V_{31} for a given set of confining pressures, the OLS method gives an accuracy of approximately 59% and 57%, respectively. Comparing it to the Random Forest method gives an accuracy of 88% and 86%, respectively (Table 8, Appendix).

For this particular sample, the RF method was found to be vastly superior to the OLS method, and the obtained values were found to be reasonably accurate when compared to those of the mathematical model.

3.3 Relationship between Geo-mechanical and Thomsen Parameters

The application of a correlation matrix interprets the relationship between the Geo-mechanical parameters and Thomsen parameters. A correlation number close to zero shows no linear link between the variables, whereas an absolute value of 1 in the correlation table indicates a complete positive linear relationship between the variables. The sign of correlation, either positive or negative, shows the direction of the relationship. If the variables are likely to decrease or increase together, then the coefficient is positive. Similarly, if one variable increases concerning a decrease in the other variable, then there is a negative correlation and coefficient if negative.

- **Dry Sandstone**

For a dry Sandstone sample, ε and γ are positively correlated to E_1, E_3, and V_{12} to a great degree, but they show a high negative correlation with V_{31}. On the contrary, δ shows the opposite trend whereby it exhibits a high negative correlation with E_1, E_3, and V_{12} whereas V_{31} is positively correlated to δ. The detailed parameters are listed in Table 9, (Appendix).

- **Shale**

For a Shale sample, ε is positively correlated to E_1, E_3, and V_{12} to a great degree, but it shows a high negative correlation with V_{31}. In the case of γ, it shows a high positive correlation with E_3 and $V_{12,}$ but it shows a high negative correlation with E_1 and V_{31}. On the contrary, δ shows the opposite trend whereby it displays a high negative correlation with E_3 and V_{12} whereas E_1 and V_{31} are positively correlated to δ. The detailed parameters are listed in Table 10, (Appendix).

- **Sandy Shale**

For a Sandy shale sample, ε and γ are positively correlated with V_{31} to a great degree, but it shows a moderately negative correlation with V_{12}. Moreover, it does not hold any relation with E_1 and E_3. On the contrary, δ shows a high correlation with all the geo-mechanical parameters. The detailed parameters are listed in Table 11, (Appendix).

- **Saturated Sandstone**

For a Saturated sandstone sample, ε is positively correlated with E_1 and V_{12} to a great degree, but it shows a high negative correlation with E_3 and V_{31}. Whereas in the case of γ, there is a stark difference as it shows a high positive correlation with E_3 and $V_{31,}$ but it exhibits a high negative correlation with V_{12}. Moreover, γ has no relation with E_1. Similarly, δ exhibits a moderate and high negative correlation with E_1 and $V_{12,}$ respectively, whereas E_3 and V_{31} are positively correlated to δ. The detailed parameters are listed in the Table 12, (Appendix).

4 Conclusion

Currently, the conventional approach of selectively adopting many technologies and applying digitalization may not be the best way forward. Instead, the industry would gain more if it pursued a transformative agenda with digitalization as the foundation itself. A digital transformation at this stage can revolutionize not only the industry but also benefit society. A centered digital strategy and a culture of creativity and technology adoption would be required for such a transition.

Through this study, we may conclude that elastic anisotropy parameters are the primary determinants of hydrocarbon reservoir characterization parameters estimation. To estimate more precise hydrocarbon reservoir characterization parameters, vertical P-wave and S-wave velocities, as well as the three anisotropy values, are required. Surface seismic data of good quality and high resolution can be used to estimate the anisotropy parameters ε, γ, and δ. We must rely on downhole data, wireline measurements for sonic profiling, and other seismic profiling methods to determine the remaining parameters. The lab studies on core samples would only aid in the development of the initial model by providing empirical connections between some of the parameters.

After applying the ML algorithms, the anisotropy parameters and the geo-mechanical properties could be estimated with reasonable accuracy. Using the mathematical model would have required us to find out the stiffness constants first, which has been eliminated using ML algorithms, which facilitate the direct estimation of geo-mechanical properties through velocity profile inputs.

Moreover, we can also conclude that for a machine learning model to predict correct values with less error margin; the model needs to be trained with more modeling data. The amount of data points we need for training a model has a substantial effect on the overall accuracy of the models. So, to be able to learn and understand the complexities, patterns, and relationships between given input and output variables, requires more modeling data.

We can understand the effect of fewer modeling data points on the overall accuracy of the OLS method, where the OLS (Ordinary Least Square) model fails miserably in predicting some data points of the original data. It means the model does not entirely understand the relationships between the variables on fewer data.

References:

[1] Das, B., Mondal, I. and Chatterjee, R., Effect of Shale Anisotropy in Modification of In-situ Stress in Krishna-Godavari Basin, India, *13th Biennial Conference and Exhibition. Kochi: Society of Petroleum Geophysicist.*, 2020.

[2] Barton, N., Quadros, E., Anisotropy is Everywhere, to See, to Measure, and to Model, *Rock Mech Rock Eng* 48, 2015, pp. 1323–1339.

[3] Schön, J. H., *Handbook of Petroleum Exploration and Production,* Elsevier, 2011, pp. 245-271.

[4] Thomsen, L., Weak Elastic Anisotropy, *Geophysics,* Vol. 52, No. 10, 1986, pp. 1954-1966.

[5] Kaarsberg, E. A., Introductory studies of natural and artificial argillaceous aggregates by sound-propagation and X-ray diffraction methods, . *J. Geol*., Vol. 67, 1958, pp. 447–472.

[6] Podio, A.L., Gregory, A. R., and Gry, K.E, Dynamic properties of dry and water saturated Gren River shale under stress, *SPE Journal*, Vol. 08, No. 04, 1968, pp. 389 – 404.

[7] Jones, L. E. A., and Wang, H. F., Ultrasonic velocities in Cretaceous shales from the Williston basin, *Geophysics,* Vol. 46, 1981, pp. 288–297.

[8] Lo, T. W., Coyner, K. B., and M. N., Experimental determination of elastic anisotropy of Berea Sandstone Chicopee Shale and Chelmsford Granite, *Geophysics*, Vol. 51, 1986, pp. 164-171.

[9] Hornby, B.E., Experimental laboratory determination of the dynamic elastic properties of wet, drained shales , *J Geophys Res.* Vol. 103, No. B12, 1998, pp. 29,945 – 29,964.

[10] Wang, Z., Seismic anisotropy in sedimentary rocks, part 1: A single-plug laboratory method . *Geophysics,* Vol. 67, 2002, pp. 1415–1422.

[11] Bjørlykke, K., Clay mineral diagenesis in sedimentary basins - a key to the prediction of rock properties. Examples from the North Sea Basin, *Clay Minerals*, Vol. 33, 1998, pp. 15 – 34.

[12] Sam, M.S. and Andrea, A., The effect of clay distribution on the Elastic properties of sandstones. *Geophysical Prospecting,* Vol. 49, 2001, pp. 128-150.

[13] Syed, F. I., Abdulla, A., Dahaghi, A. K., Neghabhan S., Application of ML & AI to

model petrophysical and geomechanical properties of shale reservoirs - A systematic literature review, *Petroleum,* Vol. 8, 2020, pp. 158-166.

[14] Rajabi, M., & Tingay, M., Applications of Intelligent Systems in Petroleum Geomechanics - Prediction of Geomechanical Properties in Different Types of Sedimentary Rocks, *International Workshop on Geomechanics and Energy: The Ground as Energy Source and Storage*, 2013, doi:http://dx.doi.org/10.3997/2214-4609.20131949.

[15] Wang, Z., Seismic Anisotropy in Sedimentary Rocks Part 1: A Single-Plug Laboratory Method. *Geophysics,* Vol. 67, No.5, 2002, pp. 1415-1422.

[16] Mavko, G., Mukerji, T., & Dvorkin, J., Tools for Seismic Analysis in Porous Media, *Rock Physics Handbook*, Cambridge University Press, 1998.

[17] Pena, F. R., Elastic Properties of Sedimentary Anisotropic. *Dissertation,* Massachusetts Institute of Technology , 1998, pp. 19-26.

[18] Jain, D., ML, R-Squared in Regression Analysis, *www.geeksforgeeks* (2023) , URL: https://www.geeksforgeeks.org/ml-r-squared-in-regression-analysis, 2023.

[19] Reinstein, I. *Random Forests, Explained*, *www.kdnuggets.com (2017)*, URL:https://www.kdnuggets.com/2017/10/random-forests-explained.html, 2017.

Appendix

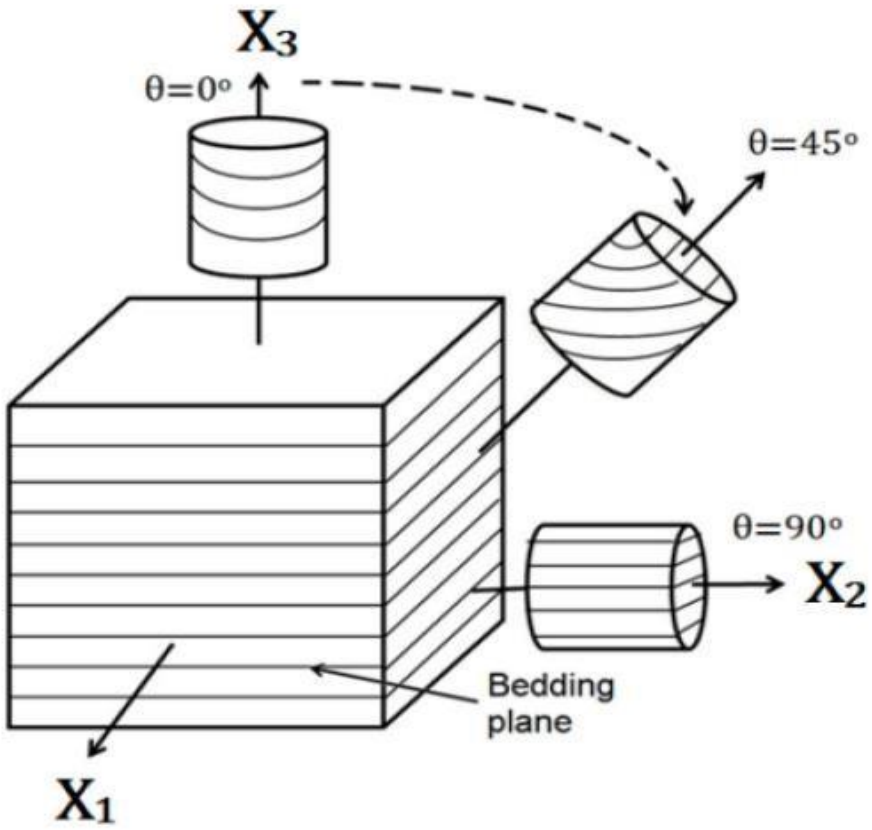

Fig. 1: Experimental procedure for data acquisition from core same, [15].

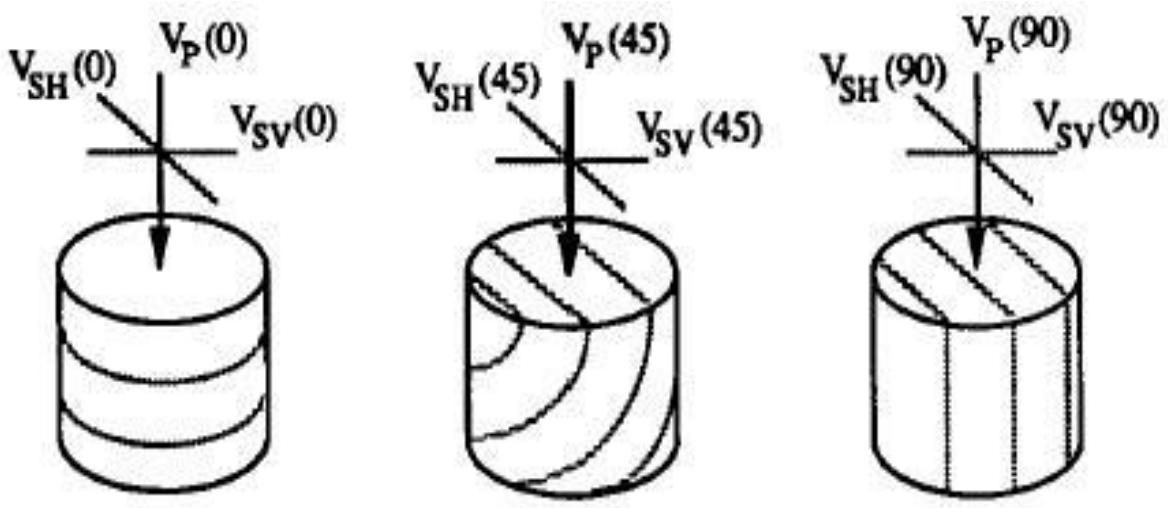

Fig. 2: Velocity measured at different angles for core samples, [15].

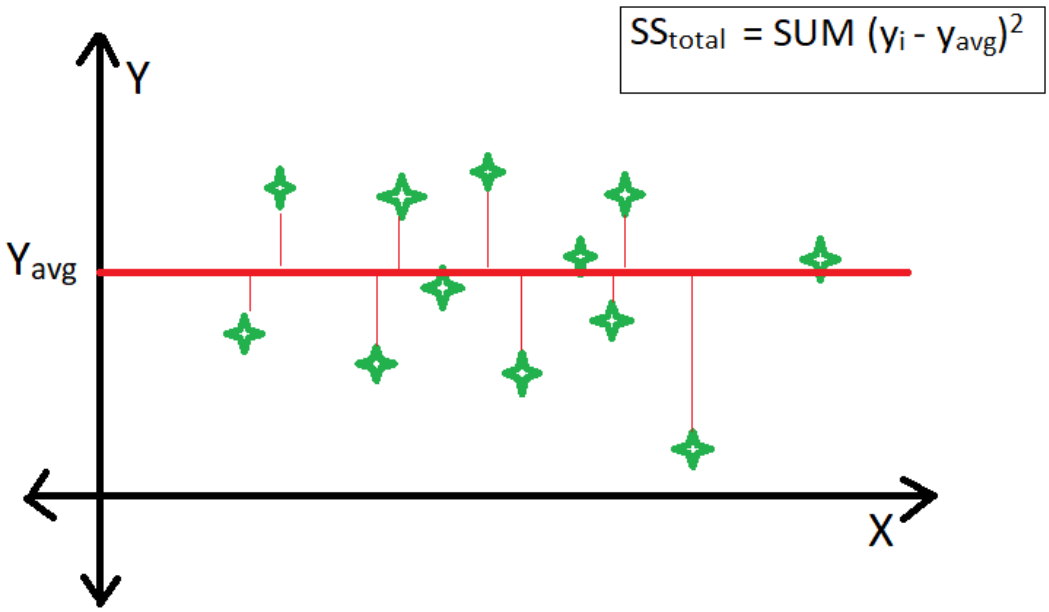

Fig. 3: Graphical representation of R^2 value, [18].

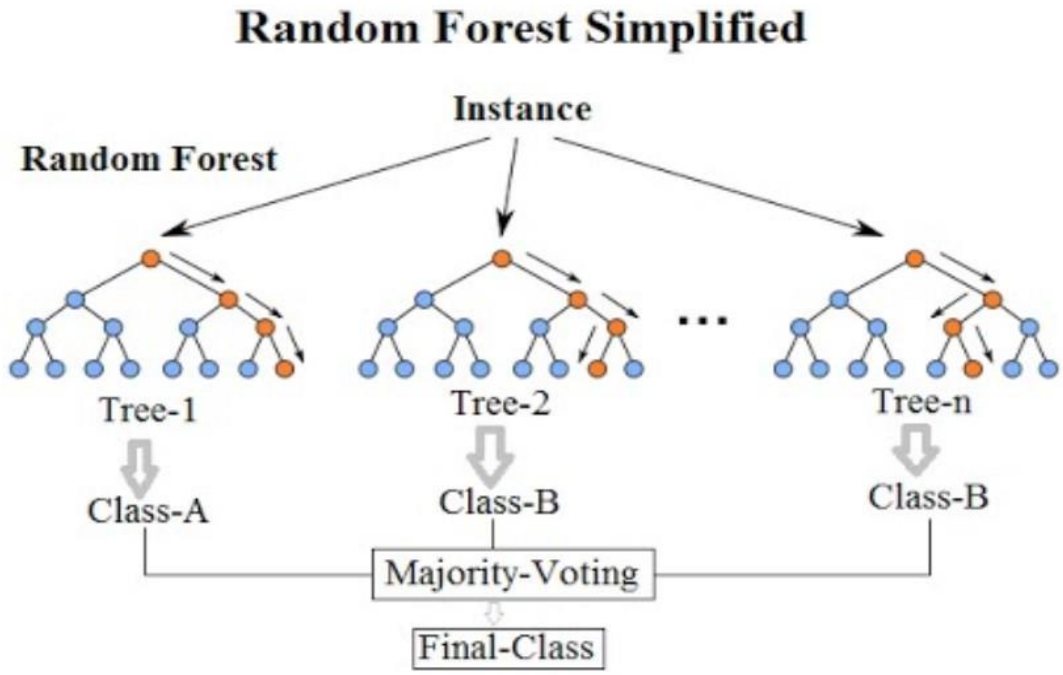

Fig. 4: Diagrammatical representation of the Random Forest method, [19].

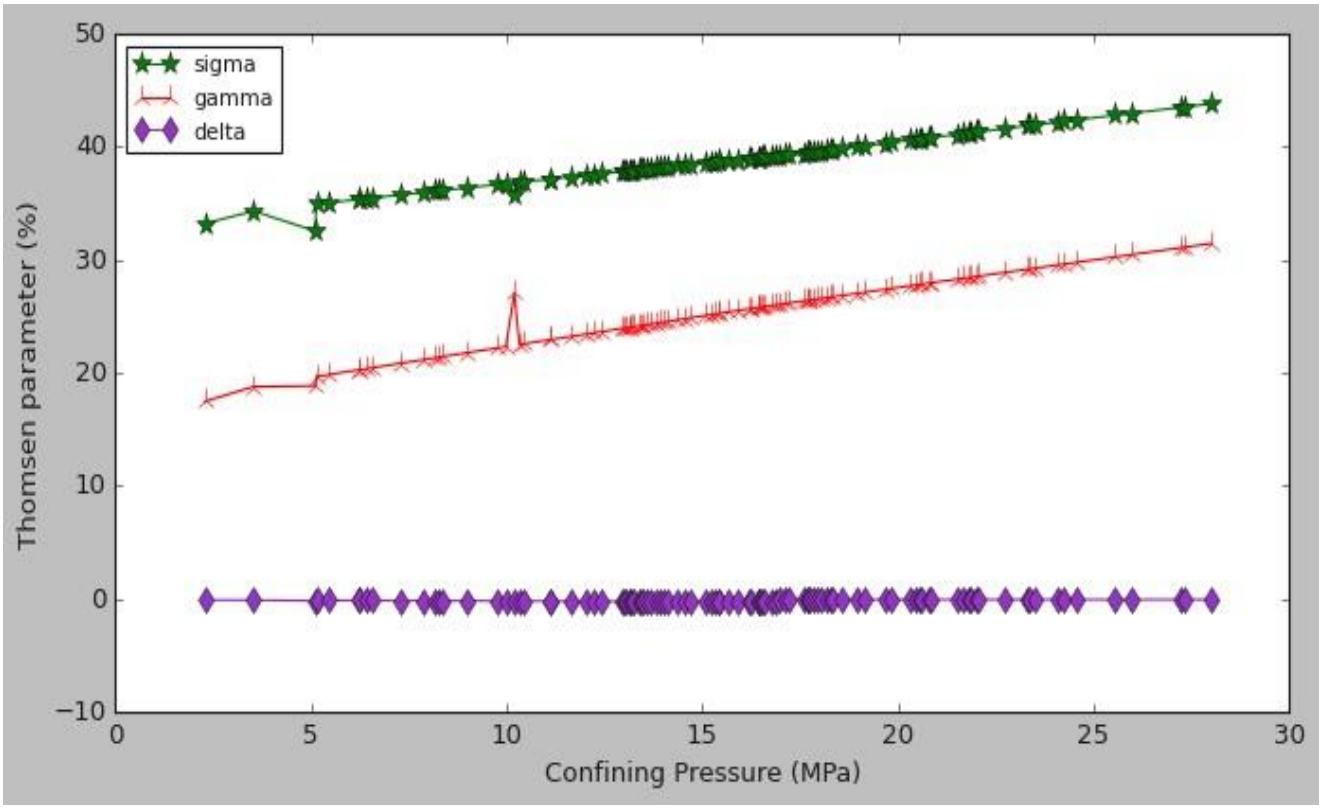

Fig. 5: Thomsen's parameters for dry sandstone as a function of confining pressure.

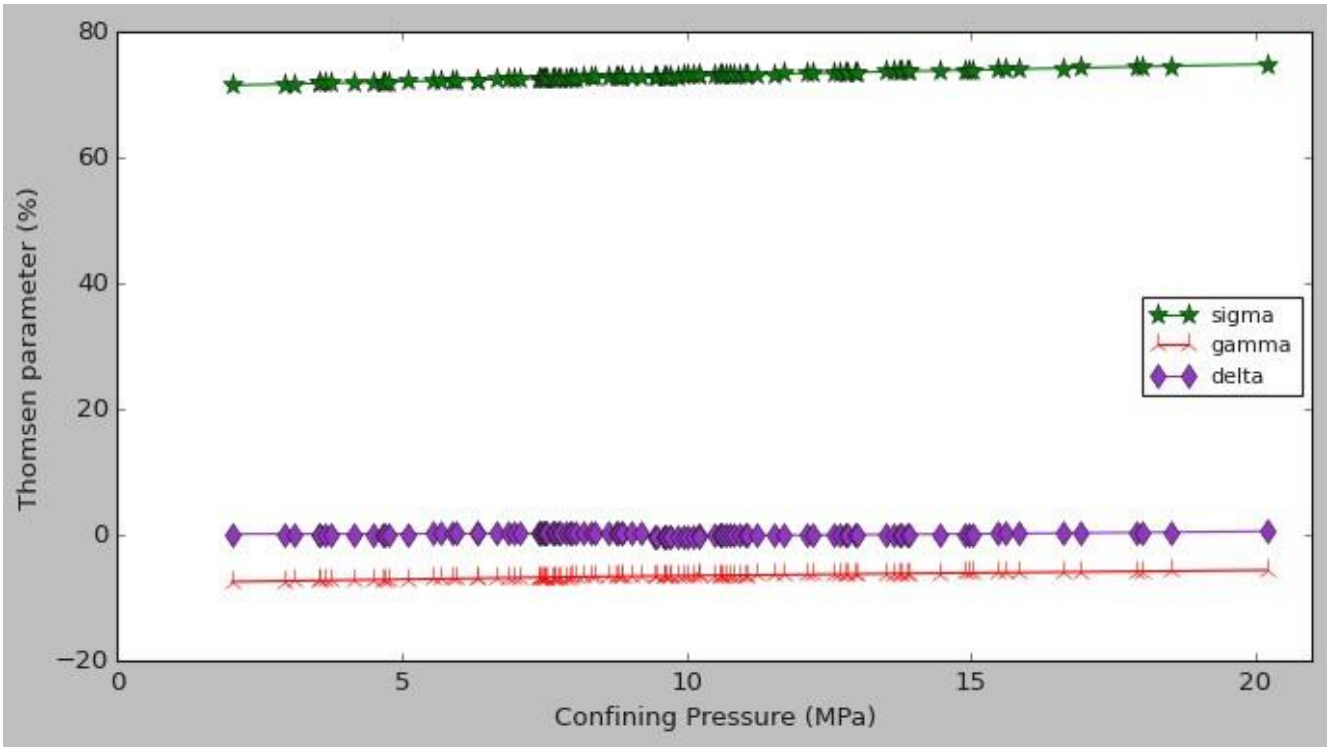

Fig. 6: Thomsen's parameters for sandy shale as a function of confining pressure.

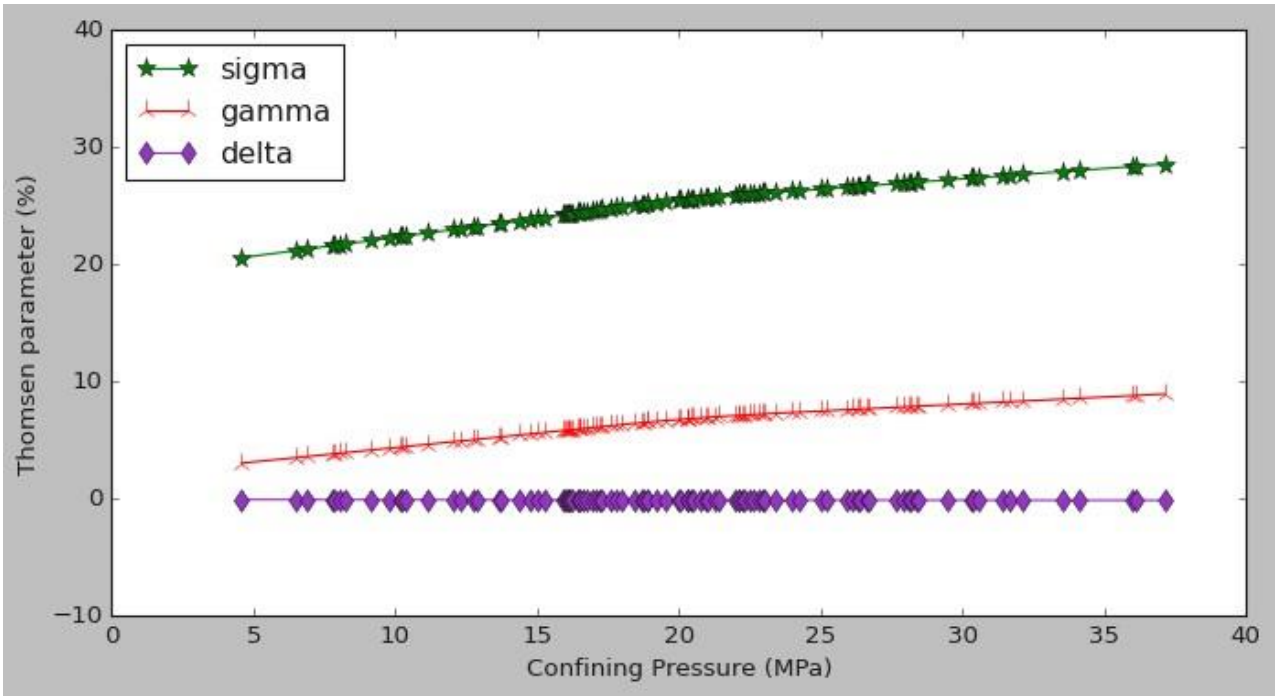

Fig. 7: Thomsen's parameters for shale as a function of confining pressure.

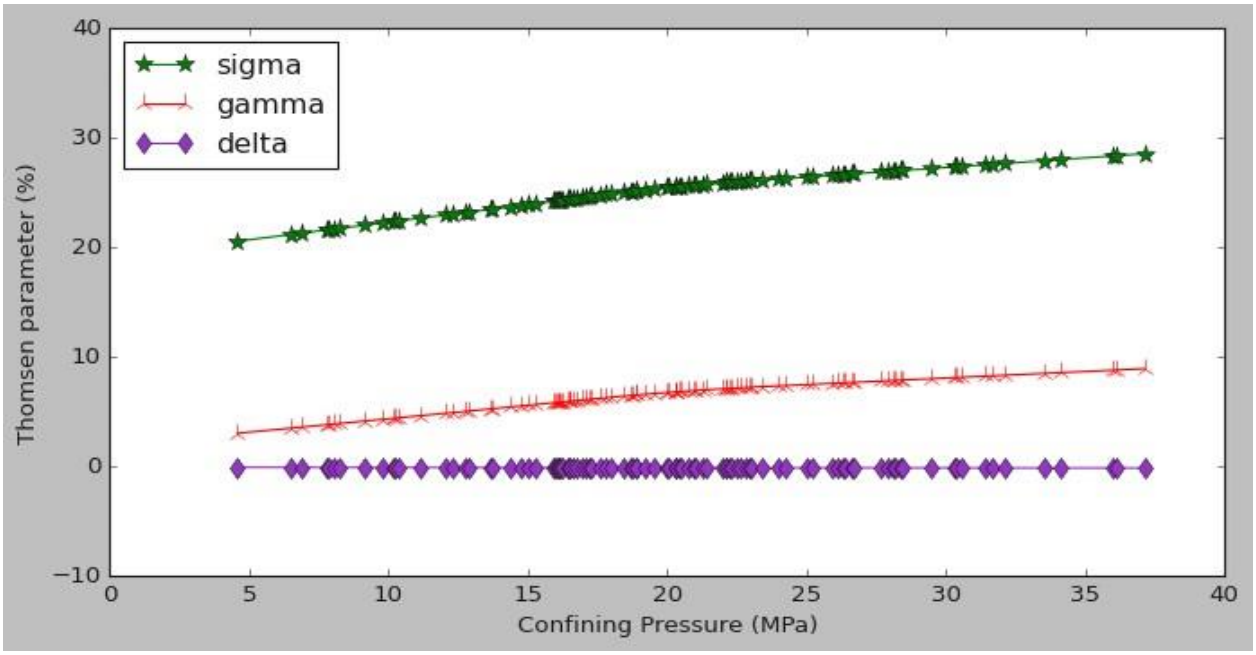

Fig. 8: Thomsen's parameters for saturated sandstone as a function of confining pressure.

Table 1. Comparison between the mathematical and predicted model for estimating Young's modulus in a dry sandstone sample.

Dry Sandstone										
Confining pressure (MPa)	**E_1 (Horizontal Young's Modulus (MPa))**					**E_3 (Vertical Young's Modulus (MPa))**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
2.3	3.708012529	3.910	95%	3.79	98%	1.410995664	1.735	81%	1.4260	99%
5.1	3.210830011	3.216	100%	4.431	72%	1.402640235	1.4	100%	1.4240	99%
10.2	3.632629248	3.668	99%	3.955	92%	1.490805682	1.494	100%	1.5137	98%
20	5.465012748			5.67	96%	1.797465001			1.7470	97%
30	6.11027495			5.95	97%	1.838503204			1.7900	97%
30.2	6.384496232			6.12	96%	1.853176489			1.8140	98%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			98%		92%			94%		98%

Table 2. Comparison between the mathematical and predicted model for estimating Poisson's ratio in a dry sandstone sample.

Dry Sandstone										
Confining pressure (MPa)	**V_{31} (Vertical Poisson's ratio)**					**V_{12} (Horizontal Poisson's ratio)**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
2.3	-0.016860547	0.0049	29%	-0.0340	50%	0.169340116	0.1780	95%	0.162	96%
5.1	-0.135471039	-0.1390	97%	-0.1220	90%	0.100161802	0.1000	100%	0.118	85%
10.2	-0.159511085	-0.1550	97%	-0.1300	81%	0.080528974	0.0810	99%	0.103	78%
20	0					0.176851374			0.189	94%
30	0					0.233715971			0.192	82%
30.2	0					0.198308787			0.195	98%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			75%		74%			98%		89%

Table 3. Comparison between the mathematical and predicted model for estimating Young's modulus in a shale sample

Shale						
Confining pressure (MPa)	**E_1 (Horizontal Young's Modulus (MPa))**			**E_3 (Vertical Young's Modulus (MPa))**		
	Mathematical Model	**Predicted Model**		**Mathematical Model**	**Predicted Model**	
		OLS Method	**Accuracy**		**OLS Method**	**Accuracy**
3.1	9.733428376	9.710	100%	6.003302606	6.010	100%
5.2	11.0755007	11.160	99%	6.158448828	6.316	98%
10.3	12.22255047	11.870	97%	6.341019392	6.390	99%
20.1	12.83280877	12.960	99%	6.503649957	6.411	99%
30.6	13.44198011	13.200	98%	6.491447506	6.594	98%
40.3	13.22874197	13.220	100%	6.620959109	6.624	100%
			Avg. Accuracy			**Avg. Accuracy**
			99%			99%

Table 4. Comparison between the mathematical and predicted model for estimating Poisson's ratio in a shale sample

Shale						
Confining pressure (MPa)	**V_{31} (Vertical Poisson's ratio)**			**V_{12} (Horizontal Poisson's ratio)**		
	Mathematical Model	**Predicted Model**		**Mathematical Model**	**Predicted Model**	
		OLS Method	**Accuracy**		**OLS Method**	**Accuracy**
3.1	0.129480576	0.129	100%	0.427600388	0.428	100%
5.2	0.104723742	0.112	94%	0.449112072	0.476	94%
10.3	0.073195914	0.068	93%	0.430831071	0.429	100%
20.1	0.051288404	0.050	97%	0.445622109	0.431	97%
30.6	0.054635724	0.054	99%	0.462438597	0.471	98%
40.3	0.058807873	0.058	99%	0.459837903	0.460	100%
			Avg. Accuracy			**Avg. Accuracy**
			97%			98%

Table 5. Comparison between the mathematical and predicted model for estimating Young's modulus in a sandy shale sample.

Sandy Shale										
Confining pressure (MPa)	**E_1 (Horizontal Young's Modulus (MPa))**					**E_3 (Vertical Young's Modulus (MPa))**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
1.3	5.566921833	5.563	100%	5.66	98%	1.32	1.31	99%	1.322	100%
3.3	5.746127921			5.405	94%	1.33			1.3	98%
5.2	5.906146999			5.583	95%	1.33			1.311	98%
10.2	5.525406037	5.536	100%	5.446	99%	1.33	1.33	100%	1.327	100%
20.2	7.043371391	7.437	95%	6.837	97%	1.36	1.43	95%	1.36	100%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			98%		97%			98%		99%

Table 6. Comparison between the mathematical and predicted model for estimating Poisson's ratio in a sandy shale sample

Sandy Shale										
Confining pressure (MPa)	**V_{31} (Vertical Poisson's ratio)**					**V_{12} (Horizontal Poisson's ratio)**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
1.3	0					0.53	0.531	100%	0.53	99%
3.3	0					0.52	-2.040	25%	0.51	98%
5.2	0					0.51	2.818	18%	0.50	99%
10.2	-0.036633466	-0.04	92%	-0.04	83%	0.49	0.490	99%	0.49	99%
20.2	0.032063781	0.05	64%	0.02	65%	0.49	0.500	97%	0.49	99%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**

Table 7. Comparison between the mathematical and predicted model for estimating Young's modulus in a saturated sandstone sample

Saturated Sandstone										
Confining pressure (MPa)	**E_1 (Horizontal Young's Modulus (MPa))**					**E_3 (Vertical Young's Modulus (MPa))**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
6.7	2.60	2.600	100%	2.562	98%	1.97	1.9660	100%	2.0270	97%
11.3	2.35	5.032	47%	2.525	93%	2.54	5.4240	47%	2.1120	83%
15.6	1.89	0.540	29%	2.502	75%	1.92	0.4087	21%	2.2071	87%
20.1	2.32	1.059	46%	2.506	92%	2.27	0.9625	42%	2.2687	100%
30.3	2.36	2.737	86%	2.523	93%	2.28	2.6780	85%	2.3032	99%
40.2	2.54	2.537	100%	2.5327	100%	2.33	2.3340	100%	2.3033	99%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			68%		92%			66%		94%

Table 8. Comparison between the mathematical and predicted model for estimating Poisson's ratio in a saturated sandstone sample

Saturated Sandstone										
Confining pressure (MPa)	**V_{31} (Vertical Poisson's ratio)**					**V_{12} (Horizontal Poisson's ratio)**				
	Mathematical Model	**Predicted Model**				**Mathematical Model**	**Predicted Model**			
		OLS Method	**Accuracy**	**RF Method**	**Accuracy**		**OLS Method**	**Accuracy**	**RF Method**	**Accuracy**
6.7	0.20	0.1950	100%	0.2131	92%	0.27	0.2650	100%	0.2534	96%
11.3	0.16	0.6700	23%	0.2365	66%	0.32	0.7870	41%	0.2390	74%
15.6	0.35	0.0460	13%	0.2641	74%	0.18	-0.0580	32%	0.2163	83%
20.1	0.30	0.0750	25%	0.2823	94%	0.21	0.0126	6%	0.1960	93%
30.3	0.32	0.3833	82%	0.2955	94%	0.16	0.2230	73%	0.1780	91%
40.2	0.31	0.3000	98%	0.2967	97%	0.16	0.1620	100%	0.1740	93%
			Avg. Accuracy		**Avg. Accuracy**			**Avg. Accuracy**		**Avg. Accuracy**
			57%		86%			59%		88%

Table 9. Correlation between Geo-mechanical and Thomsen parameters for a dry sandstone sample

Dry Sandstone				
	E_1	**E_3**	**V_{12}**	**V_{31}**
ε	0.9997	0.9987	0.9994	-0.9999
γ	0.9987	0.9972	0.9985	-0.9998
δ	-0.9998	-0.9997	-0.9988	0.9989

Table 10. Correlation between Geo-mechanical and Thomsen parameters for a shale sample

Shale				
	E_1	**E_3**	**V_{12}**	**V_{31}**
ε	0.9997	0.9987	0.9994	-0.9999
γ	-0.9997	0.9972	0.9985	-0.9998
δ	0.9996	-0.9997	-0.9988	0.9989

Table 11. Correlation between Geo-mechanical and Thomsen parameters for a sandy shale sample

Sandy Shale				
	E_1	**E_3**	**V_{12}**	**V_{31}**
ε	-0.0844	0.0111	-0.5469	0.9844
γ	-0.1241	-0.0308	-0.5821	0.985
δ	0.985	0.8983	0.8018	0.9704

Table 12. Correlation between geo-mechanical and Thomsen parameters for a saturated sandstone sample

Saturated Sandstone				
	E_1	**E_3**	**V_{12}**	**V_{31}**
ε	0.8515	-0.9906	0.9307	-0.985
γ	-0.1388	0.7323	-0.8713	0.7551
δ	-0.5919	0.9665	-0.997	0.9764

Contribution of Individual Authors to the Creation of a Scientific Article:
Dr. Jwngsar Brahma is responsible for overall supervision, writing the original draft, the writing review, editing Normal analysis, Investigation, Validation, writing of the original draft and the writing review.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself:
No funding for this research work.

Conflict of Interest
The author has no conflict of interest to declare.