Journal of Oil, Gas and Petrochemical Technology 10(1): 1-24, Winter and Spring 2023

RESEARCH PAPER

Young's Modulus Estimation Using Machine Learning Methods and Daily Drilling Reports

Parirokh Ebrahimi[1,2], Ali Ranjbar[1,2*], Fatemeh Mohammadinia[1,2], Hojat Ghimatgar[3], Abbas Hashemizadeh[4]

[1] *Faculty of Petroleum, Gas and Petrochemical Engineering, Persian Gulf University, Bushehr, Iran*
[2] *Sustainable Exploitation of Underground Resources Research Group, Persian Gulf University, Bushehr, Iran*
[3] *Department of Electrical Engineering, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr, Iran*
[4] *Department of Chemical Engineering, University of Qom, Qom, Iran*

ARTICLE INFO

Article History:
Received 19 August 2022
Revised 9 January 2023
Accepted 6 February 2023

Keywords:
Daily Drilling Report
Young Modulus
Geomechanical Parameters
Support Vector Machine
Artificial Neural Network
Random Forest
LSBoost
Baysian

ABSTRACT

To avoid drilling damages, it is very important to determine the field stress. Prediction of elastic parameters such as Poisson's ratio and Young's modulus is of great importance in determining in-situ stress and completing geomechanical modeling. These parameters are calculated statically through laboratory tests on drilling cores or dynamically through log data. However, such data may not be available in the oil field data-bank. Therefore, Daily Drilling Reports (DDR) can be introduced as a suitable alternative for predicting rock elastic modulus. In this study, for the first time, an attempt was made to estimate the Dynamic Young's modulus using DDR data with the application of a variety of conventional machine learning methods. In this regard, linear, support vector machine (SVM), artificial neural network (ANN), Random Forest (RF) LSBoost, and Baysian have been used. Input data to these algorithms also include depth, string rotary speed (RPM), rate of penetration (ROP), weight on bit (WOB), density (RHOB), porosity (Φ), pump pressure (PP), and tangential velocity (TV). Each of these algorithms was then compared in terms of accuracy using correlation coefficient (R2), mean squared error (MSE), and root mean square error (RMSE) criteria. Finally, using conventional experimental correlations and core data, the resulting values were converted to static values. The results showed that using daily drilling reports, based on the above criteria, a good estimate of the elastic parameters can be achieved. Moreover, Baysian and LSBoost methods had slightly higher and better accuracy than other methods.

How to cite this article
Ebrahimi P, Ranjbar A, Mohammadinia F, Ghimatgar H, Hashemizadeh A., Young's Modulus Estimation Using Machine Learning Methods and Daily Drilling Reports, Journal of Oil, Gas and Petrochemical Technology, 2023; 10(1): 1-24. DOI:10.22034/jogpt.2023.356733.1107

1. INTRODUCTION

In the upstream technology of the petroleum industry, especially in wellbore drilling operations, many mechanical problems and challenges exist, and solving them requires understanding the geomechanical behavior of the hydrocarbon reservoirs. In this regard, calculating the elastic and strength parameters of the rock and also

* *Corresponding Author Email: Ali.ranjbar@pgu.ac.ir*

interpreting the in-situ stresses are of great importance. These parameters can be effective in preparing geomechanical models [1, 2]. Elastic parameters have a significant impact on several aspects in the oil industry such as wellbore stability [3-7], determination of safe mud window [8, 9], well completion [10, 11], exploitation [12], reservoir evaluation operations [13, 14] and field development [15].

In other words, preventing possible problems caused by not understanding the stress in the well, such as pipe stuck in the well, falling of the well, and loss of the well, requires the calculation and understanding of the stress in the well. The stress calculation is obtained from the integration of geomechanical parameters. The most important geomechanical properties of the rock are Young's modulus (*E*) and Poisson ratio (*v*) which define the relationship between strains and stresses that create deformation [16]. The Poisson ratio is an elastic geomechanical property that is the ratio between transverse and axial strains and is defined according to Equation 1. The Poisson's ratio for isotropic and elastic rocks is linearly between 0 and 1, and for most rocks is between 0 and 0.5. Young's modulus indicates the rock's stiffness, defined by the relationship between stress and strain according to Hooke's law (Equation 2) [17].

$$\varepsilon_{trans} = v\varepsilon_{axial} \tag{1}$$

$$\sigma = E\varepsilon_{axial} \tag{2}$$

where ε_{trans} and $\varepsilon_{\mathrm{axial}}$ are represented by transverse and axial strain, respectively. *v* is the Poisson ratio, σ is axial stress, and *E* is Young's modulus. These elastic properties can be predicted directly through compression tests on core plug samples in a laboratory by applying axial load or from petrophysical and acoustic well loggings (using bulk density, and compressive and shear wave velocity) [18, 19]. In the first case, the loading and unloading speed is significantly slower than in the second one. Therefore, the elastic parameters calculated in the laboratory are generally defined as "static elastic" parameters, and the elastic parameters calculated by petrophysical logs are defined as "dynamic elastic" parameters [20-22]. In contrast with dynamic elastic parameters, static ones play a key role in geomechanical studies and interpretations. Due to the lack of a sufficient number of cores along the well and the cost of laboratory tests, it is not possible to continuously obtain static geomechanical characteristics; therefore, dynamic measurements are used in this regard. In other words, dynamic geomechanical data obtained from well loggings can be calibrated with static data. To determine the dynamic Poisson ratio (v_{dyne}), and dynamic Young's modulus (E_{dyne})from sonic well log data, Equations 3 and 4 are used [23].

$$v_{dyne} = \frac{V_p^2 - 2V_s^2}{2(V_p^2 - V_s^2)} \tag{3}$$

$$E_{dyne} = \rho V_s^2 \frac{3V_p^2 - 4V_s^2}{V_p^2 - V_s^2} \tag{4}$$

where V_p and V_s are compressive and shear wave velocities obtained from DSI (Dipole Sonic Imager) well log. and *p* is density. The relationships between dynamic and static parameters have been studied since the early 19th century when sound waves propagation techniques were used to describe rocks in mining, oil, and geotechnical engineering [20, 24]. There are different methods for measuring static values. Many researchers have used theoretical and experimental research to provide empirical correlations to estimate the static properties from the dynamic ones. Table 1 shows well-known empirical correlations presented for the different types of rock lithology by different researchers. However, many of these empirical correlations have been predicted using limited samples for a specific type of formation, and their use in different fields will face many uncertainties. Instead of using the mentioned relations, core data and its correlation to petrophysical logs can be used. As mentioned above, this requires the availability of logging data to calculate dynamic elastic parameters through mathematical equations (e.g., Equations 3 and 4). However, logging is a generally costly and time-consuming operation. In addition, for many reasons such as indirect measurements, limitations caused by tool specification, environment effects, varying resolution, sufficient log data may not be available within the target reservoir thickness. Therefore, mud logging unit can be used as alternative data with high potentials to estimate the dynamic elastic parameters of the rock using artificial intelligence and machine learning methods [25-29].

Table 1. Empirical correlations for predicting the static Young's Modulus

Author	Correlation	Input	Data sets	Lithology	R^2
King [30]	$E_{st} = 1.263E_{dyne} - 29.5$	E_{dyne}	174	Igneous and Metamorphic	0.82
Najibi et al [31]	$E_{st} = 0.014E_{dyne}^{1.96}$ $E_{st} = 0.169V_P^{3.324}$	E_{dyne}, V_P	45	Limestone	0.87 0.90
Eissa and Kazi [20]	$E_{st} = 0.74E_{dyne} - 0.82$	E_{dyne}	76	Sedimentary	0.96
Van Heerden [32]	$E_{st} = aE_{dyne}^{b}$	E_{dyne}	14	Different sandstones, quartzites, norite, and magnetite	0.95
Fei et al [18]	$E_{st} = 0.564E_{dyne} - 3.4941$	E_{dyne}	22	tight sandstones	0.7246
Per Horsrud [33]	$E_{st} = 0.169V_P^{3.324}$	V_P	NA	Shale	0.9
Christaras et al [34]	$E_{st} = 3.02e^{0.00055V_P}$ $E_{st} = 1.05E_{dyne} - 3.16$	E_{dyne}, V_P	8	gypsum, phonolite, basalts, granite, limestone, and andesite	0.97 0.99
Feng et al [35]	$E_{st} = 0.81E_{dyne} - 13.88$	E_{dyne}	18	sandstone and siltstone rocks	0.84

Machine learning concepts. Machine learning has become one of the important tools in engineering and industry. It is used in predicting, diagnosing, and describing trends from various aspects [36-38]. Today, machine learning is often used for general reservoir description, core, log, and drilling data processing [39]. According to the existing studies, the most important issues that have been analyzed based on machine learning techniques include:

- Casing collapse risk assessment and depth prediction [40],
- Models development for rock strength [41],
- Acoustic logs prediction (P wave and S wave) [42, 43],
- Rock failure parameters prediction [44],
- Establish a relationship for bubble point pressure [45],
- Rate of penetration prediction and optimization (ROP) [46-49],
- Static elastic modules prediction [50],
- Fracture and pore pressure prediction [51],
- Total organic carbon prediction [47, 52-55],
- Gas saturation in organic formations prediction [56],
- Formation Lithology identification [57, 58],
- Formations tops (anticline) prediction [57, 59],
- New hybrid hole cleaning model [60].

One of the most prominent studies in this field is the use of drilling data in estimating various important parameters. For example, a machine learning model has been introduced by Abdelgawad et al. [61] for determining rheological properties based on mud density, Marsh funnel viscosity (MFV), and solid percent. The support vector machine model developed by Ahmed S, A et al [62], using drilling parameters such as weight on bit (WOB), round per minute (RPM), rate of penetration (ROP), mud weight (MW), Torque (TORG) as the model input parameter and pore pressure as the target parameter. Finally, the developed model based on the Gaussian Kernel function, with an accuracy of R^2 over 0.995, predicted the pore and fracture pressure. In 2020, Ahmed et al [63], by using 5 machine learning methods including functional networks (FN), artificial neural networks (ANN), support vector machine (SVM), radial basis function (RBF), and fuzzy logic (FL) estimated the fracture pressure and compared it with the empirical correlations in this regard. Onshore drilling data include D, HKHT, HKL, FPWPMP, ROP, SPP, TOR, WOB, RPM, W_{OUT}, CON_{IN}, TMP_{IN}, MW_{IN}, Pp, P_F. These data were used in the construction of the ANN model with R^2=0.99 and AAPE=0.094% to estimate the pressure of fracture in this study. Among these researches, the study by Abdelaal et al [64] can be mentioned. The scholars used machine learning and drilling data to estimate the pore pressure.

In 2017, datasets from three wells in Alberta, Canada were used by Atashnezhad et al [65], to construct an experimental model to estimate porosity. To find the best model, a differential evolution (DE) algorithm was applied. In this study, the developed model of unlimited compressive strength (UCS) obtained from gamma rays at the bit (GR) and the model of penetration rate (ROP) which is a function of drilling parameters such as WOB, RPM, bit geometry, hydraulic drilling, and rock formation were used to estimate the porosity of the formation. Finally, a model was proposed in which the porosity could be generated from drilling data in addition to the UCS value.

In 2018, ROP estimation was conducted by Diaz et al [66, 67], with ANN modeling and decision regression, using drilling data such as depth, pore pressure gradient, equivalent circulating mud density at the bottom hole, rotary speed, Reynolds number, the fraction of the bit tooth worn away, weight on bit by its diameter. In this way, three different scenarios of training data were evaluated and 5% of the future section was predicted. The first trained scenario used cumulative data, the second one used different amounts of data, and the third performed square root resampling. The results showed that the total average values of mean percentage error for accumulated training, fixed sizes of 300, 200, and 100 points, and data square root resampling were 21%, 24.5%, 29.2%, 20.5%, and 15.6%, respectively. In general, accumulative training captured the general ROP trend with depth, but its accuracy decreased in deeper sections. Data resampling returned the lowest mean percentage errors. Finally, it was concluded that cumulative data and data resampling could be used during ROP forecasting to assist in drilling management and well planning.

Machine Learning Application in Estimating Elastic Modulus. As mentioned in section 1, the use of experimental methods in the calculation of elastic parameters has one or more shortcomings, for example, they are valid only at a certain depth range and only for a certain lithology. Empirical constants are different from field to field. Moreover, these correlations are predicted using limited parameters whereas it is obvious that other effective parameters like rock density, porosity, UCS, etc. can play important roles. Therefore, along with the increasing progress of science and efficiency of machine learning methods and the lack of limitations mentioned above, the machine learning methods can be used as a suitable alternative with appropriate estimation accuracy. Due to the lack of sufficient core data and the high cost of coring as well as the possibility of unavailability of petrophysical log data, the use of drilling data seems a good alternative. It is because these data are always available at the lowest cost. From a practical point of view, data such as drilling depth, WOB, ROP, SPP, RPM, T, and Q are applicative parameters that are obtained during drilling [68]. Therefore, these parameters can also be related to the elastic parameters, which depend on the behavior and properties of the reservoir rock. In recent years, the use of drilling data to predict some essential parameters for future safe drilling has been developed [28, 29, 69]. To develop models for estimating critical parameters, the use of machine learning algorithms along with these data sets has a significant application in various aspects of the oil industry. Several illustrious and related studies can be mentioned as follows. In 2017, in a study by Asadi et al. [70], porosity, bulk density, acoustics logs, and ROP drilling data were used as four input parameters for the artificial neural network, and UCS was used as the target output of the network; 77 core samples datasets were used to develop an artificial neural network model; Among these, 60% of the datasets were used for model training, 20% for validation and the 20% remaining for testing. Finally, with the help of a designed network, which had 3 hidden layers and 10 neurons, a model was developed for estimating UCS with a correlation coefficient of 1 (= 1). Many researchers have also predicted the rock resistance parameter using machine learning and drilling data. Relevant studies include Amani et al. [71], Kalantari et al. [72], and Gamal et al. In 2021, Young's modulus was predicted by Elkatatny [29] based on drilling data (WOB, ROP, torque, SPP, drill pipe rotation speed, Q) with the development of an artificial neural network model. He trained 2054 data from well A and tested and validated the model on 871 and 2912 data from well B and well C, respectively. The developed ANN model with errors of 0.986, 0.981, and 0983, for training, testing, and validation respectively predicted . In 2021, Siddig et al predicted the Poisson ratio with the drilling data mentioned above and used machine learning algorithms with an accuracy of more than 0.98 and 0.97, for ANN and adaptive neuro-fuzzy inference system (ANFIS), respectively. [28]. In addition, many researchers have predicted the Poisson ratio using machine learning methods and drilling data in recent years [69, 73]. Table 2 summarizes the recent research in this regard. Based on the follow-up of previous studies, it can be found that the important parameters such as pore pressure, porosity, uniaxial strength of reservoir rock, Poisson ratio and Young's modulus, and so on can be predicted using machine learning methods. Using modern machine learning methods saves time. A review of previous studies showed that estimating the elastic parameters of rock such as Static Young's modulus is an up-to-date topic and demands further research to use drilling data for estimation.

Table 2. Models developed for elastic and rock strength parameters using AI

Author	Output geomechanical parameter	method	Inputs	Data sets	Lithology	R^2
Mahmoud et al [74]	E_S	FNN	V_S,V_P, RHOB	592	Sandstone	0.99
Al-anazi et al [75]	E_S	ACE	V_S, V_P, RHOB, P_P, depth, overburden stresses,S_h, porosity	602	NA	0.99
Mahmoud et al [50]	E_S	ANN with an optimization algorithm	V_S,V_P, RHOB	592	Sandstone	0.99
Asadi [70]	UCS	ANN	Bulk density, Sonic travel time, Formation Porosity, ROP	77	sandstone	1
Elkatatny [29]	E_{dyne}	ANN	WOB, ROP, T, Q SPP, pipe rotation speed,	2054	sand, shale, and carbonate	0.98
Siddig [28]	v_{dyne}	ANN, ANFIS	WOB, T, SPP, Q, ROP	2905	carbonate, sandstone, and shale	0.99 0.98
Ahmed [69]	V_S	FN, RF	WOB, T, SPP, RPM, ROP, Q	1775	shale, sand, and carbonate	0.89 0.94
Gamal et al [76]	UCS	FN, RF based on PCA	WOB, RS, T, SPP, Q, ROP	2333	carbonate, sandstone, and shale	0.97 0.99
Siddig [77]	E_S	ANN, SVM	ROP, WOB, SPP, T, Drilling fluid pumping rate, Rotary speed RPM	2280	sandstone, shale, and limestone	0.92 0.89
Abdulraheem et al [78]	E_S	NN, FN, FL	DT (Compression wave travel time, RHOB	77	NA	0.73 0.52 0.52
Abdulraheem et al [78]	V_S	NN, FN, FL	DT (Compression wave travel time), RHOB	77	NA	0.40 0.34 0.30
Zeeshan Tariq et al [79]	V_S	FN	V_S ،V_P، Gamma-ray, bulk density, porosity	580	Carbonate	0.97
Tariq et al [80]	V_S	ANN	V_S, V_P	550	Carbonate	0.98
Mingming He [81]	UCS	DCNN	thrust force, torque, penetration rate, rotation speed	More than 30 case	Silurian marble, sandstone, limestone, granite, shale, phyllite, rock grain, gneiss, quartzite, and slate	0.99

In this study, we attempted to estimate Dynamic Young's modulus using Daily Drilling Reports (DDR) instead of logging or mud-logging unit data. The used DDR data include depth, string rotary speed (RPM), rate of penetration (ROP), weight on bit (WOB), density (RHOB), porosity (Φ), pump pressure (PP), and tangential velocity (TV). Since the purpose of this study was to comprehensively investigate this issue, most of the conventional machine learning methods were evaluated, which adds to the novelty of this article. In this regard, a variety of machine learning methods including Linear, ANN, SVM, LSBoost, Baysian, and Random Forest (RF) were used. The following sections discuss the way the models are compared with each other.

Furthermore, after estimating Young's modulus along the well, the values of the static Young's modulus were calculated using core data. Next, the results were compared with the values obtained from conventional experimental methods.

2. Case Study

The studied well is located in one of the oil fields in the Zagros sedimentary basin in Bushehr province, located in southwest of Iran. The Zagros basin includes the sedimentary sequence and geological layers of the Cambrian to the present day, which contains the giant oil and gas reservoirs of Iran and the world [82]. The folded and trusted Zagros belt is the result of the convergent movement of two Iranian and Arabian tectonic plates that began in the Laramie phase of the Mesozoic [83] and the late Cretaceous [84]. This convergent movement has caused huge faults, folds, and other geological structures such as diapirs. The Zagros Basin is bounded on the east and southeast by the Minab and Zendan faults. In the west, this basin extends from the western boundaries of Iran to Iraq and beyond. It is bounded on the north and northwest by the Great Zagros, and on the south by the Persian Gulf and Saudi Arabia with gentle or no folds. Based on underground geology and structural geology studies, the Zagros Basin is divided into several sub-zones (Figure 1) [85].

The studied field is located in the north and east of Bushehr and in terms of geological divisions in the border between coastal Fars and the folded part of the southern Dezful embayment. Among the important anticlines of these areas that have been affected by the Kazerun-Qatar fault are Giskan, Siah, Chahpir, Kaki, Khormoj, and Kuh Namak anticlines. In this study, drilling data, core interpretations, petrophysical logs, and a geomechanical model of one of the wells in the field, which is an exploration well, were analyzed. The purpose of drilling this well was to study and evaluate Kangan and Dalan formations in terms of reservoir potentials. Figure 2 shows the stratigraphic sequence of the formations of the study area.

The large carbonate limestones of Dalan and Kangan are the main reservoirs in the southern fields of Iran. Based on age studies, the age of these formations has been reported as Late Permian and Early Triassic, respectively. Dalann Formation is located between the clastic formations of Faraghan at the bottom and Kanganate carbonate at the top [88, 89]. This formation consists of three parts, the upper and lower parts which are carbonate, and the middle part which is evaporative sediments. Kangan Formation consists of carbonate sequences (limestone, dolomitic limestone, and dolomite) along with anhydrite interlayers. In general, the presence of various facies and various diagenetic processes has caused the reservoir quality in these structures to be different in various parts.

3. Methodology

In this section, first, how data sets were prepared are explained. Then the theory of each machine learning method is presented. In majority of studies, the maximum value of several different performances is reported as the final result. However, due to the performance of machine learning algorithms, which are highly dependent on the training stage, the better results in a particular training mode are not necessarily an evidence of proper performance. In addition, when there is not enough data for training, it is not possible to rely on better performance with one run of the algorithm.

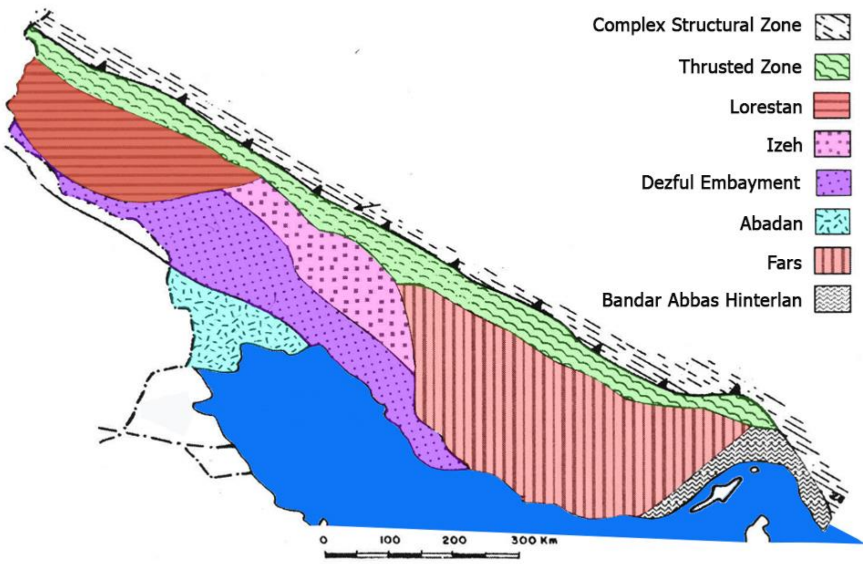

Figure 1. Subzones major of the Zagros [86]

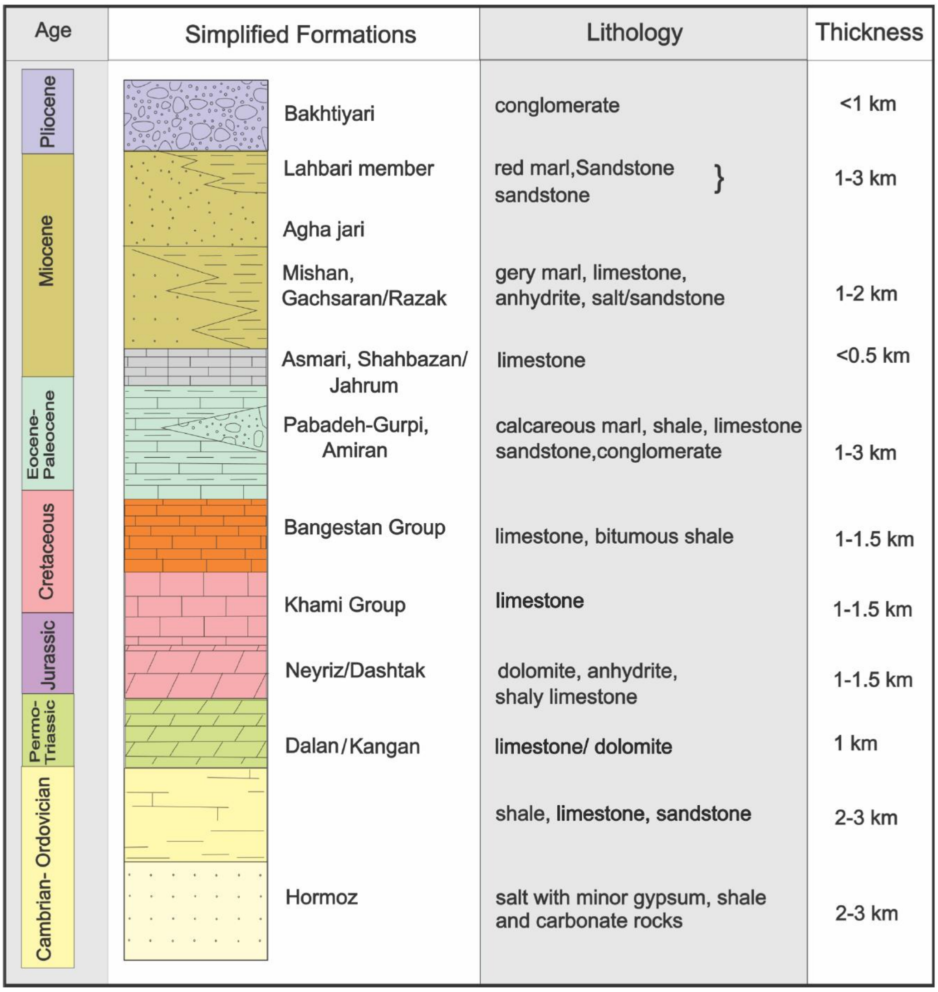

Figure 2. Stratigraphic column of formations below Fars's region, Zagros basin [87]

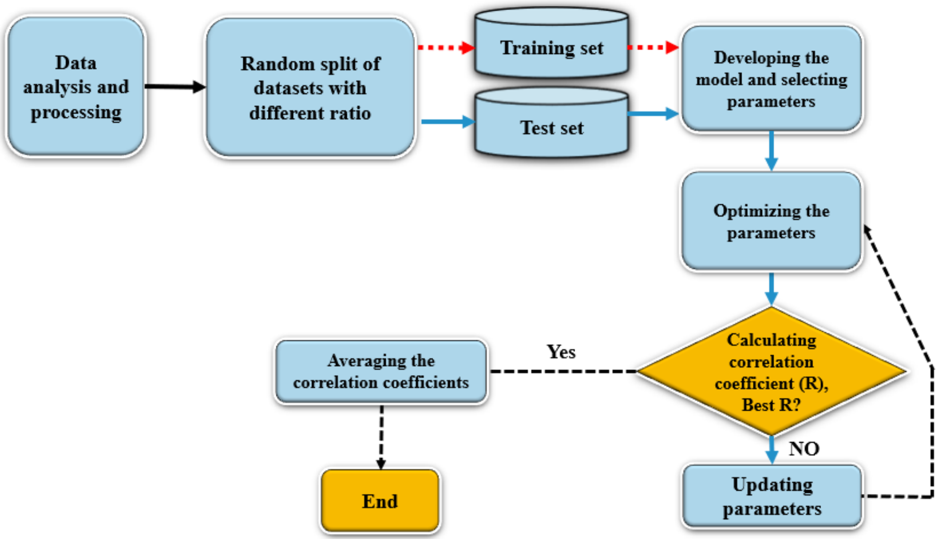

Figure 3. Flowchart of machine learning modeling

Therefore, unlike other studies, in this article averaging was performed and reported on 20 performances. In this case, the validation of the results is more valid. The flowchart of the proposed algorithm in the two stages of training and testing is shown in the figure below (Figure 3).

3.1. Data Sets and their preparation

To begin the process, daily drilling reports related to one of the wells studied in the reservoir formations (Kangan and lower Dalan Fm.) were available as image files. All daily reports were carefully reviewed and a complete set of required data was stored in a database. The datasets used in this study, which were measured during drilling at the surface, were depth, ROP, WOB, RPM, RHOB, Φ, pump pressure (PP), and tangential velocity (TV). These data were used as input parameters to estimate dynamic Young's modulus. The dynamic Young's modulus is calculated via sonic well log data according to Equation 4 mentioned in section 1 and is given as output data to machine learning models. The statistical analysis of the given datasets to determine and distinguish the statistical descriptors are presented in Table 3. Statistical analysis shows that DDR reflects good data representation and distribution. As a result, all drilling parameters were selected as input data to obtain the desired range of results.

Table 3. Statistical analysis of drilling datasets

Parameters	ROP (ft/hr)	WOB (Klb_f)	PP (PSI)	TV (Lit/M)	RPM (Es.)
Mean	2.202	19505.95	713.571	134.452	145.143
Min	0.58064	5000	400	95	58
Max	5.714	30000	1200	300	215
Mode	1.5	25000	550	128	60
Median	2.202	20000	625	128	149
Range	5.134	25000	800	205	157
Standard Deviation	1.089	6347.608	224.786	30.781	45.307
Skewness	0.709	-0.34783	0.811	2.0711	-0.294
Kurtosis	0.258	-0.684	-0.456	8.683	-0.839

3.2. Artificial Neural Network (ANN)

ANNs are one of the most widely used machine learning techniques. ANN is a computational and information processing system that mimics the behavior of biological systems and the functional characteristics of neural networks such as human and animal brains. Each artificial neural network, regardless of the implementation structure, has several basic features that are mentioned below [90]:

- An artificial neural network contains a large number of processing elements called "neurons".
- All of these processing elements have a large number of weighted connections between them.
- Relationships between elements represent the distribution of the data presented.
- A learning process is implemented to acquire knowledge.

Figure 4 shows the simple structure of various types of artificial neural networks.

ANN is interconnected based on a set of artificial neurons [91, 92]. Processing units are usually arranged in a multi-layer topology that is configured as an input layer, one or more hidden layers, and an output layer. Each unit consists of four basic components: input data, weights, a transfer function, and output values. The input for each neuron is multiplied by the corresponding weight (W_1j, W_2j, W_3j, W_4j…, W_nj). The collection value is then passed to a transfer function (for example, tan-sigmoid "TANSIG", log-sigmoid "LOGSIG"), which is applied to the weight inputs of a neuron to produce the final output [92]. Multilayer perceptron (MLP), briefly, is the classic type of neural network. They consist of at least three node layers: an input, a hidden, and an output layer. The data is first passed to the input layer, then applied to one or more hidden layers, and finally prospects are done in the output layer (also called the visible layer). MLP uses a supervised learning method called backpropagation for training [91]. Multiple layers and its nonlinear activation function distinguish MLP from linear perceptron, in which case it can well separate data that cannot be separated linearly [79].

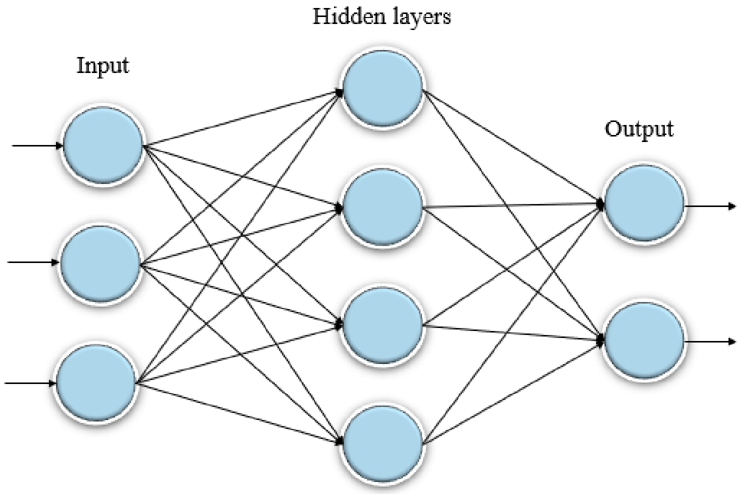

Figure 4. Schematic of the simple structure of the four-layer ANN network [90]

Because the MLPs are a function of the global approximation shown in Cybenko's theory [93, 94], they can be used to create mathematical models by regression analysis. MLPs are very beneficial in their research because of their ability to solve very complex problems randomly, which often allows approximate solutions. MLPs were a common solution for machine learning in the 1980s, but then faced stiff competition from simple support vector machines [95, 96]. In literature, many successful applications of ANN in the petroleum industry have been reported [49, 73, 97-99].

3.3. Least Square Boosting (LSBoost)

The Boosting method uses a combination of groups of models (e.g., decision tree) to find solutions to classification and regression problems. The performance of the models alone is poor, while when their predictions are combined, a strong model with high performance (high accuracy in classification problems and low prediction error in regression problems) is obtained [100]. The Boosting method focuses on reducing the bias and improving the model fit with the data. The Least Squares Boosting Algorithm is a regression Boosting algorithm that uses least squares as a cost function to fit regression sets by reducing the mean squared error. At each step, a new learner receives training with a difference between the actual value and the estimated value by all the trainees which were previously created. Figure 5 shows the steps for the LS Boost algorithm.

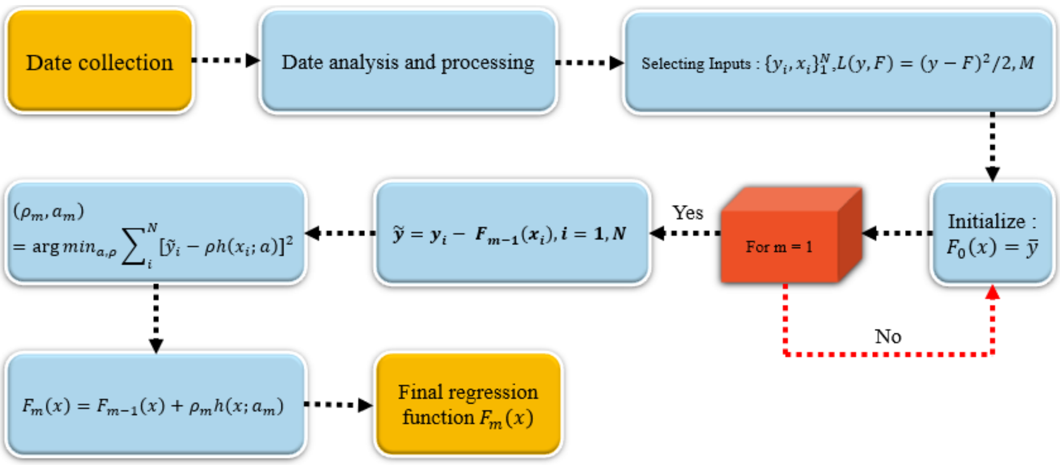

Figure 5. Steps of LSBoost algorithm

In this figure, $F_0(x)$ is the initial guess function, F_m is the incremental function, p_m is the weight of the m-th model, a_m is the division variable and *h (x; a)* is the basic learner function. *x* and *y* represent the input and output variables, respectively. $[X_i, Y_i]^N$ represent the training data, M represents the number of iterations, and *L* represents the cost function which is defined as follows [101]:

$$L(y,F) = \frac{(y-F)^2}{2} \tag{5}$$

The proposed model is obtained by assigning different weights to each tree to minimize the mean squared error between the actual value of y and the estimated value of $\tilde{y}$. Initially, the model uses the average of all $\tilde{y}$ target values and the trees add weight regression to minimize the mean squared error as follows:

$$\tilde{y}(x) = \tilde{y}(x) + \vartheta \sum_{n=1}^{N} w_n T_n(x) \tag{6}$$

Where $\tilde{y}(x)$ represents the predicted values, W_n is the weight of the nth learning tree, T is the weak learning tree, and ϑ is the learning rate, which ranges from 0 to 1.

3.4. Baysian Network

The history of the Baysian network goes back to the discovery of the Baysian formula discovered by the English priest Thomas Bayes in 1763. Later the Baysian network was first designed by Judae Pearl in 1988 [102]. Important advantages of this model include the management of lost values between input data, risk analysis and uncertainty with more accuracy than other models, providing approximate solutions using simulation techniques or prediction methods in cases where an exact solution is not available, and the possibility of combining quantitative and qualitative data. The Baysian network includes a probabilities chart model that indicates a set of random variables and their independent probabilities associated with each other. The structure of a Baysian network is a graphical representation network that is graphically modeled from the vague relationships of the variables. In addition to showing the quantity of the relationship between the variables as a numerical distribution that uses their common probability, it also shows the quality of the relationship between the problem variables. This network is a straightforward and cyclical diagram in which nodes are the variables of the problem. The following equation shows the Baysian correlation [103, 104]:

$$P\ (C_i\ |X) = \frac{P\ (X\ |C_i)\ P\ (C_i)}{P\ (X)} \tag{7}$$

Where is *P(Ci)* the probability of occurrence *C, P (X)* is the probability of occurrence *X, P(Ci | X)* is the probability of occurrence *C* occurs if *X* occurs and $P\ (X\ |\ C_i)$ is a probability of occurrence X occurs if *C* occurs.

Each Baysian network consists of three main components, including a set of nodes, a set of clauses, and a set of probabilities [102]. In general, nodes are either parents or children. A child node can be generated by multiple parents. Nodes that are preceded by another node in the graph are defined by a conditional probability distribution. Otherwise, they are expressed with the initial probability. The probabilities related to the lowest part of the Baysian network are obtained through the law of total probability, and the probabilities related to the upper parts of the network are based on the Baysian law. As mentioned above, Baysian networks allow the analyst to make forward and backward calculations. In fact, not only from the aggregation of the status of the causal parameters to the disabled status can be achieved, but also in this method, by having the disabled status or the same parameter predicted by a backward process, it will be possible to calculate the status of effective parameters. In other words, it can be determined how much each parameter will affect the final risk. Of course, it is clear that if the Baysian network model is developed and analyzed based on very little observational information, the error of analysis and related predictions will increase.

3.5. Support Vector Machine (SVM)

The support vector machine was first introduced by Vapnik in the 1990s. It is an efficient learning system and one of the supervised methods which is used for the classification and regression. Support vector machine system uses the inductive theorem of structural error minimization to provide an overall optimal solution[105]. Similar to all regression problems, in the SVM regression model the target is to determine the relationship between input and output variables with an algebraic function such as *f(X)* plus the allowable error value ε. This relationship is a function related to the dependent variable *Y*, which itself is a function of several *X* variables.

$$f(x) = W^T.\emptyset(X) + b \tag{8}$$

Where W^T is the transpose coefficients vector and b is the constant of the regression function properties and also $\emptyset$ is the kernel function. In this case, the target is to find the function *f (X)*. By training the SVM model, the objective function is determined by a set of data [106]. To determine W and b, minimizing the error function is necessary, Equation (9), in the SVM model by considering the limitations, Equation (10) [107].

$$min\frac{1}{2}W^TW + C\sum_{i=1}^{N}(\varepsilon_i + \varepsilon_i^*) \tag{9}$$

$$\begin{cases} y_i - (W^T.\emptyset(X) + b) \leq \ \varepsilon + \varepsilon_i \\ (W^T.\emptyset(X) + b) - y_i \leq \ \varepsilon + \varepsilon_i^* \\ \varepsilon_i , \varepsilon_i^* \geq 0 \ , \ \ i = 1,2, \dots N \end{cases} \tag{10}$$

In the above equations, *W* is the weight vector, and *C* is the positive real constant, which determines the penalty when a model training error occurs. $\emptyset$ is a kernel function, *N* is the number of data, and the two properties $\varepsilon_i , \varepsilon_i^*$ are slack variables. In this algorithm, a kernel function is used to solve the problem of large-scale operations, and the input vectors are plotted into a multidimensional space. A hyperplane is then created to separate the input vectors as much as possible. In this case, the operation can be performed with the same speed as the input data space. In other words, by using the kernel function, the problem of multidimensional and nonlinear plot is solved. The good performance of this algorithm depends on the good choice of parameters and kernel function. Figure 6 shows the structure of the support vector machine [108].The input data of the SVM model in this study are the drilling data mentioned in the previous sections (including depth, string rotary speed (RPM), rate of penetration (ROP), weight on bit (WOB), density (RHOB), porosity (Φ), pump pressure (PP), tangential velocity (TV) and the output is the Young Dynamic Modulus. This model processes the input data by using Kernel function. Then by connecting the output and input data, it is ready to estimate each new input data.

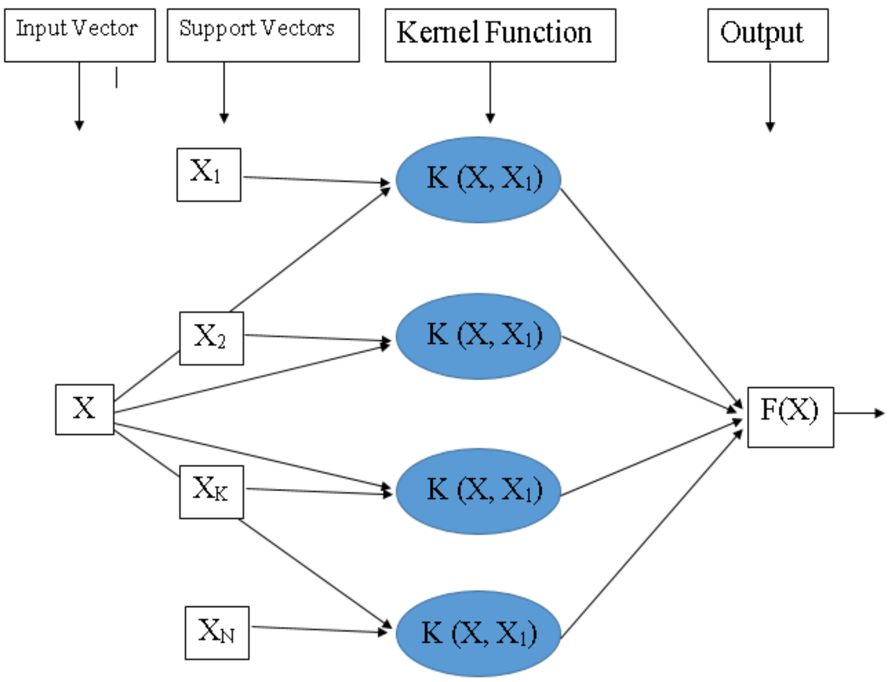

Figure 6. Support vector machine structure

3.6. Random Forest (RF)

Random Forest is a predictor and learning model consisting of a set of N random regression trees and can be used to solve classification, regression, and other problems by constructing a large number of decision trees during training [109]. Random Forest divides the data set into training and testing or OOB (out of a bag) for classification tasks. During the training phase, Random Forest uses hundreds of sets of decision trees, each containing thousands of decision leaves [110-112], and for regression tasks, the average or mean prediction of individual trees go back [113]. This computational model proposes a way to extend training and testing datasets by creating tree classifiers with unlimited capacity while increasing accuracy for both datasets. Random Forest modifies tree decisions to adapt to the dataset; however, Random Forest combines many classifiers to provide solutions for complex problems [109, 114]. Figure 7 shows the RF regression algorithm schematically.

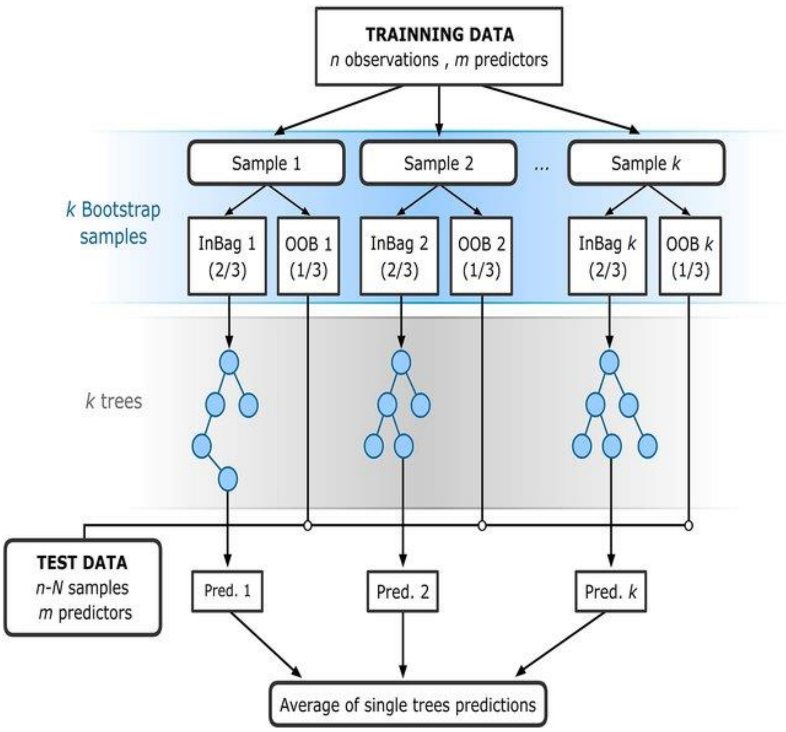

Figure 7. RF regression algorithm [115]

4. Results and Discussion

In this study, in line with previous studies, six machine learning approaches (namely SVM, ANN, LS Boosting, Baysian, Linear, and RF) were used to estimate Young's modulus. In this section, first, the parameters of each forecast model are set. Each model has one or more different parameters that the performance of each model can be affected by changing the relevant parameters. Also, the effect of training to test data (the ratio of 70% training data and 30% test data) for optimized models was evaluated to obtain the optimal value. The division of these two sets is described in detail in section 4.2. In the following, the overall performance of different algorithms for the set parameters is evaluated based on three criteria: RMSE, MSE, and correlation coefficient (R^2) for 20 different implementations with different training sets. The implementation of regression algorithms and performance evaluation of these algorithms were conducted using MATLAB software on a system with 8 GHz RAM and 2.7 GHz GPU.

4.1. Adjusting the optimal parameters of each model

To determine the best parameters of each model in the algorithms, different runs were performed in each method. In this way, each of the six machine learning methods was evaluated in several different runs for a wide range of adjustment parameters of each model. In the ANN algorithm, the number of layers, layer sizes, activation functions, and step size tolerance were considered as regulatory parameters. Although the evaluated algorithms had many regulatory parameters, the parameters that significantly affected the performance of the regression algorithm, were focused on. Regulatory parameters in SVM were considered as kernel type and optimization process, in RF as several trees and type of prediction, and in LSBoost as several training courses and learning rate. The results of optimizing the SVM parameters in Figure 8 are shown as boxplot diagrams according to R^2 criterion for 20 different performances. The length of each plot box represents the scatter of values around the mean, and the red line represents the mean of the values per different runs. In the SVM algorithm, the best performance was obtained for linear kernel and optimization routine as SMO (Social Media Optimization). It is also observed that by choosing the optimization routine as ISDA (Intelligent Systems Designs and Applications), the results are much lower.

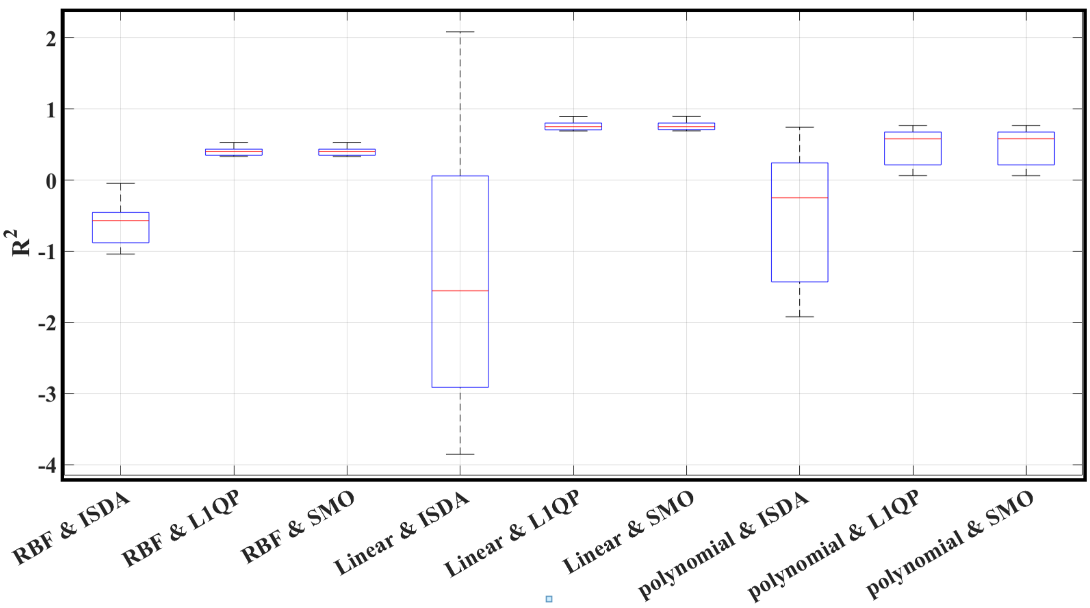

Figure 8. Adjust SVM parameters

The optimization results of the RF algorithm parameters are shown in Figure 9. Values in Figure 9 are given for different quantities of Nt = 10, 20, …, 100 and for two cases with / without OOB Prediction on 20 runs. As can be seen in most cases (except Nt = 50), the results obtained with OOB prediction are much better. Also, for the value of Nt = 60, the highest performance of the RF algorithm is obtained. Therefore, the number of trees is equal to 60 and the prediction process of the OOB prediction type is selected. Figure 10 shows the results of the LSBoost algorithm for two different parameters. The values of learning cycles vary from 10 to 100 with step size = 10, and the values of the learning rate are shown in different colors for the values 0.1, 0.4, …, and . As can be seen, for less learn rate (Lr), the algorithm gives better results. In addition, for learning cycles and learn rate, 50 and 0.1better algorithm results were obtained, respectively (higher mean and lower standard deviation). In the Baysian method, the lasso regression model, which prior distributions conditioned on σ^2 was used. Each regression coefficient in this distribution is double exponential values and they are priori and independent. The mean of these values is zero and their scale is σ/λ . In this relation λ is the lasso shrinkage parameter that its increase will lead the coefficients to tend towards 0. B and σ^2 are dependent random variables, respectively.

Table 4 summarizes the optimized parameters with the lowest prediction error and the best performance of each algorithm.

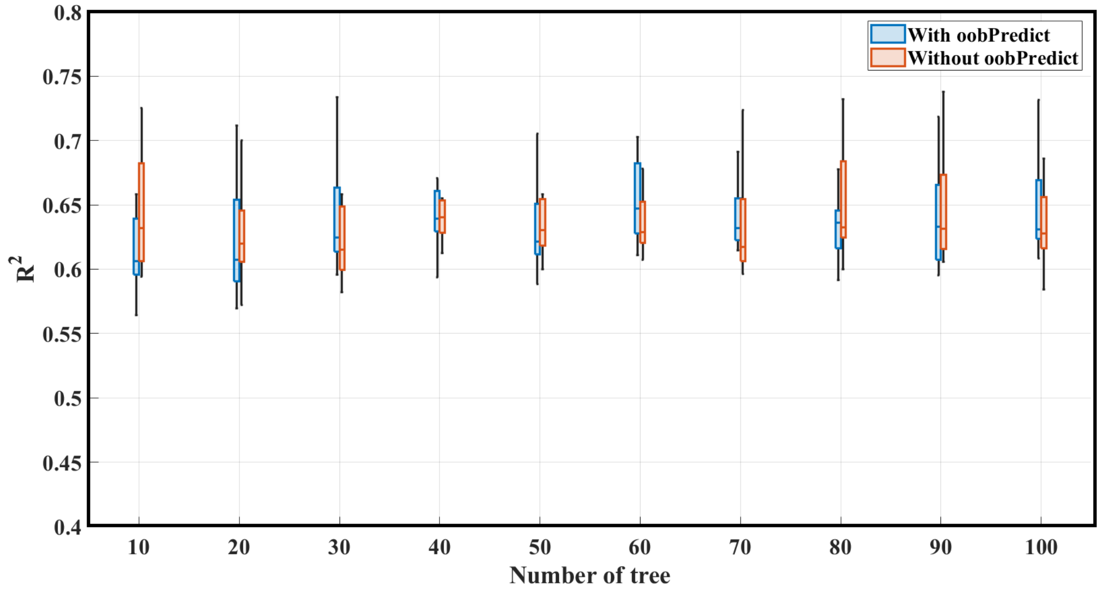

Figure 9. Plot box RF performance for different number of trees and two prediction modes with / or without OOB Prediction

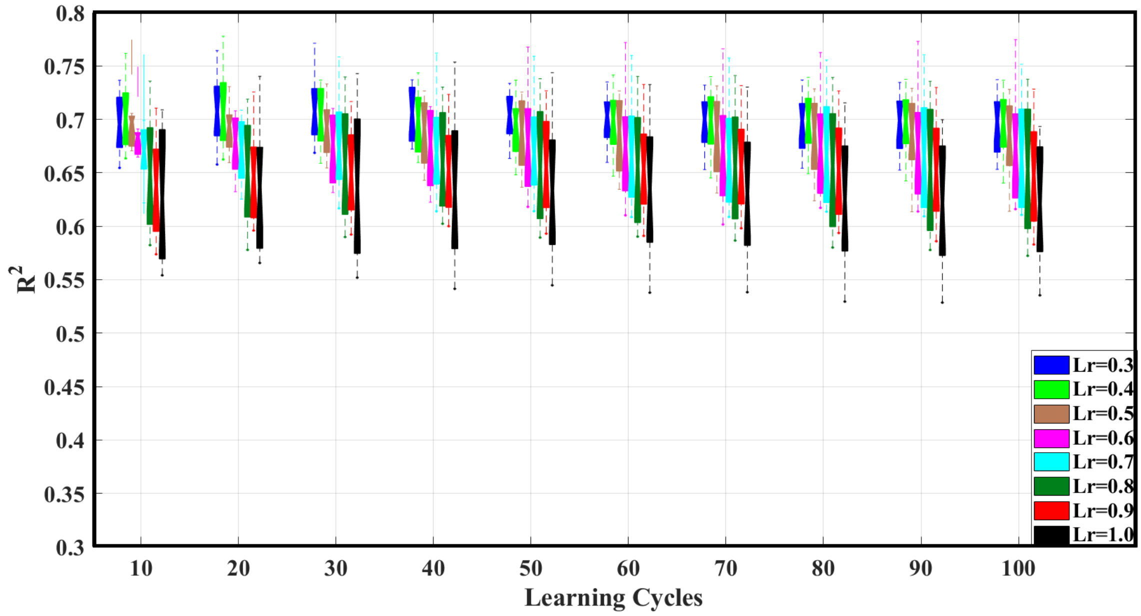

Figure 10. The correlation coefficients versus learning cycles of 20 performance in LSBoost algorithm results

Table 4. Design parameters of optimal models for estimating E_{dyne}

NN	
Number of layers	2
Size of layers	[20 5]
Activation function	Relu
Step size tolerance	10^{-6}
SVM	
Kernel	Linear
Optimization routine	Iterative Single Data Algorithm
RF	
Number of trees	50
Type of prediction	OOB Prediction (out-of-bag prediction)
LSBoost	
learning cycles	50
Learn Rate	0.1

4.2. Adjust training data ratio

The ratio of training data used in the training phase to the total data used in the study can greatly affect system performance. A low training data ratio leads to not recognizing the existing patterns in the problem and reduces the performance of the algorithm. Conversely, if the ratio of training data to total data (as well as test data) is very high, overfitting usually occurs (meaning that the accuracy of the algorithm is very good in the training suite and far in the test suite). Figure 11 shows the performance of various algorithms for the ratio of training data to the total available data set (α = 10%, 20%, ..., 90%). With the increase of training datasets, the performance of almost all evaluated algorithms up to 70% is always improved. From 70% onwards, in all algorithms except the NN algorithm, the performance of the other algorithms either decreases or does not change much. In the following results, the educational data ratio is considered to be 70% and the test data ratio is considered to be 30% of the total available data.

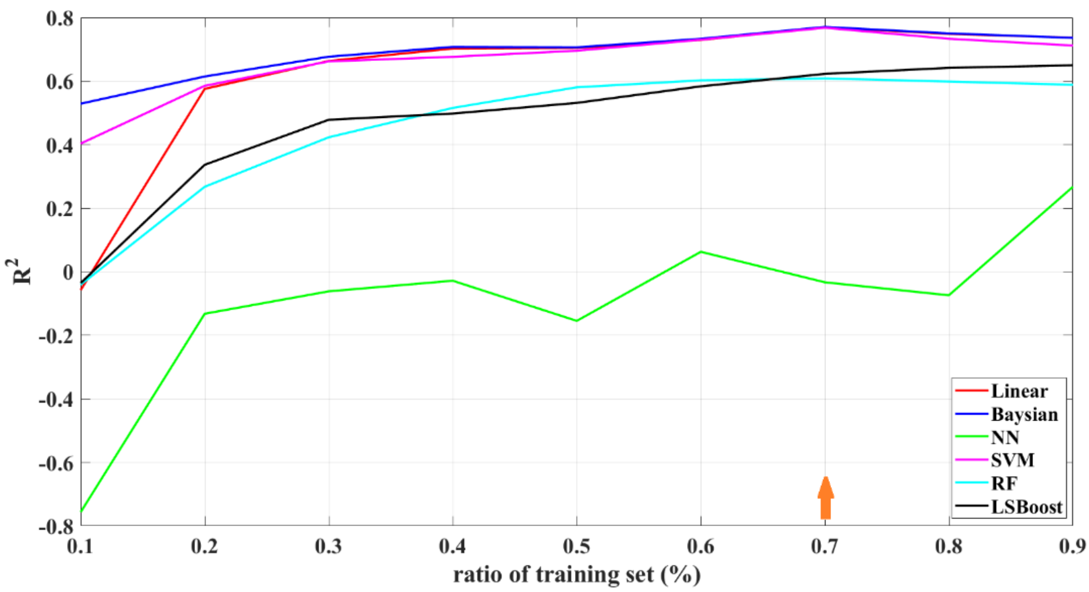

Figure 11. Evaluating and determining the ratio of training and testing

4.3. Regression Results

The results of this section are based on the optimized parameters in Table 4 as well as the ratio of 70% training data and 30% test data (based on the results of Figure 11). The results are presented for both training and test data sections based on different evaluation criteria in Table 5. In this table, the mean and standard deviation of RMSE, R^2 and MSE values for each of the studied algorithms are reported for 20 different runs and two sets of training and testing separately. The performance of the NN algorithm on the training set is significantly better than other algorithms; however, this performance is not very reliable because the performance of this algorithm per test set, is much worse than other algorithms (average performance 0.08 for R^2). According to the training test set, the performance of the Baysian algorithm with an average RMSE of 7.34, an average R^2 of 0.76, and an average MSE of 54.15 are better than the other algorithms evaluated. Moreover, the standard deviation of the values obtained from the Baysian algorithm in different runs is the lowest value compared to other algorithms.

Table 5. Mean and standard deviation of Young's module estimation results using Linear, Baysian, NN, SVM, RF and LSBoost algorithms based on three evaluation criteria RMSE, R^2and MSE on 20 different performances

	RMSE		R2		MSE	
	Train	Test	Train	Test	Train	Test
Linear	6.35±0.17	7.40±0.61	0.79±0.02	0.75±0.05	40.34±2.20	55.14±9.19
Baysian	6.35±0.17	7.34±0.57	0.79±0.02	0.76±0.05	40.41±2.20	54.15±8.54
NN	1.26±3.44	20.43±5.37	1.00±0.00	0.08±0.30	12.64±47.46	444.33±232.10
SVM	6.70±0.17	7.32±0.53	0.77±0.02	0.75±0.04	44.91±2.30	53.80±8.05
RF	6.05±0.16	9.43±0.93	0.81±0.01	0.67±0.05	36.59±1.90	89.75±18.10
LSBoost	3.06±0.17	8.84±0.50	0.96±0.01	0.63±0.05	9.42±1.07	78.34±9.10

In Figure 12 to Figure 17, the results of different algorithms in their best runs are given for two sets of tests and training. Since the performance of the algorithm also depends on the training and the test set, the best run from the 20 performances of each algorithm is selected as follows:

$$x = max\ (R_{test}^2 \times R_{train}^2) \tag{11}$$

That is, the run is selected as the best performer that maximizes the result by multiplying and . The LSBoost algorithm has the highest performance in the test set and training set. After the LSBoost algorithm, Baysian, Linear, and RF algorithms have the highest performance. Based on the previous results and the results of this section, it can be claimed that the two algorithms Baysian and LSBoost offer the highest performance.

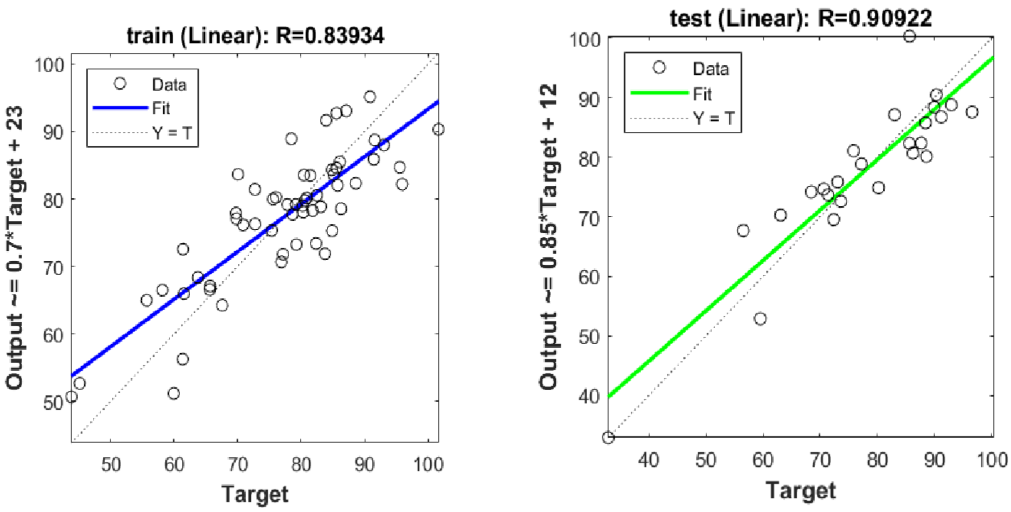

Figure 12. Results of different algorithms for both train and test datasets in the linear network

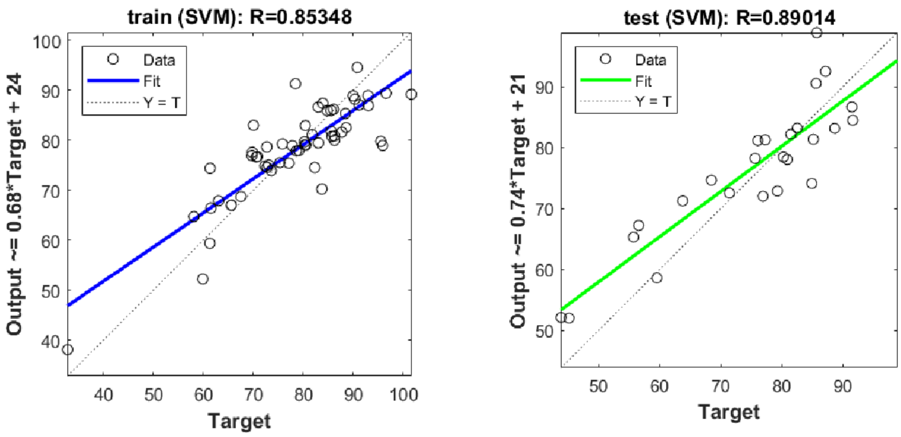

Figure 13. Results of different algorithms for both train and test datasets in the SVM network

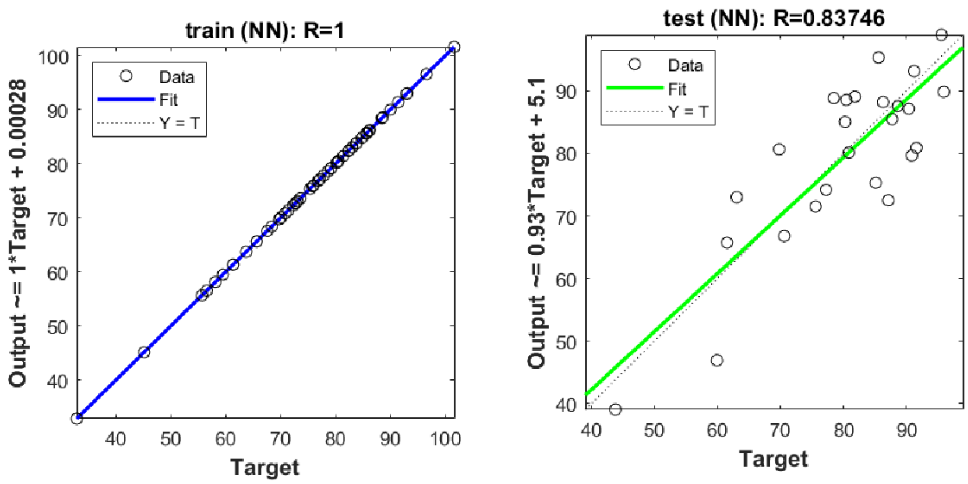

Figure 14. Results of different algorithms for both train and test datasets in the NN network

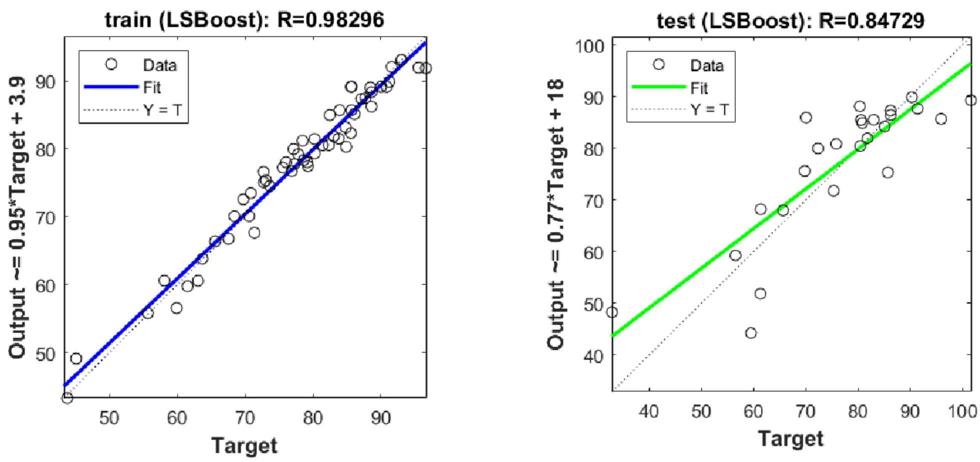

Figure 15. Results of different algorithms for both train and test datasets in the LSBoost network

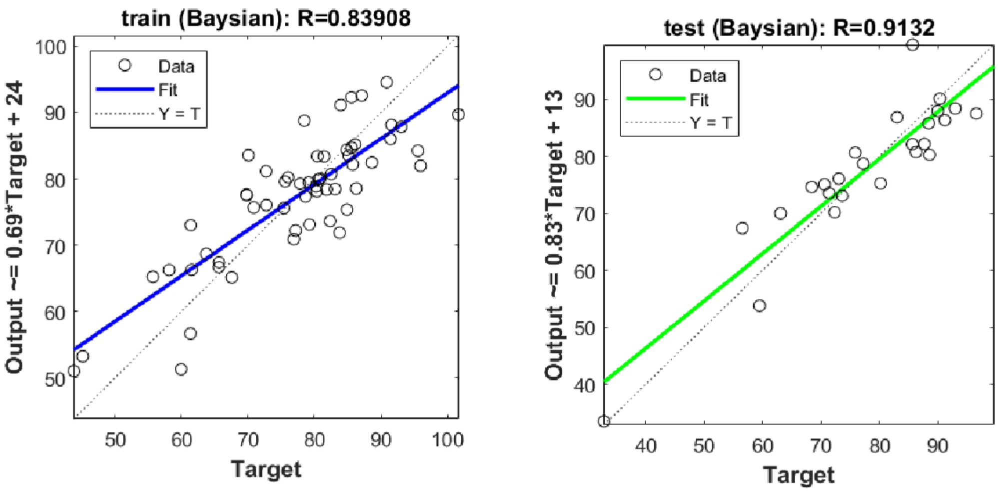

Figure 16. Results of different algorithms for both train and test dataset in Baysian network

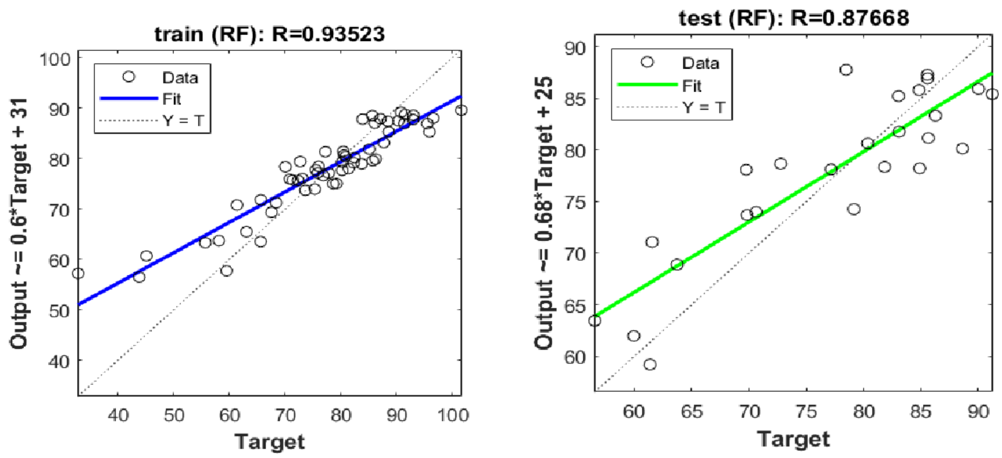

Figure 17. Results of different algorithms for both train and test dataset in RF network

4.4. Converting Dynamic to Static Young's Modulus

After predicting Young's dynamic modulus using daily drilling reports, as a geomechanical parameter, it needs to be calibrated with static data obtained from core data (laboratory results). In section 1, it was mentioned that many experimental correlations were proposed by researchers to predict the static Young's modulus in Table 1. In this paper, some mentioned experimental correlations of static Young's modulus including King, Eissa and Kazi, Feng et al, Fei et al, and Christaras et al were used to calibrate the dynamic Young's modulus. Comparisons obtained from the calibrated Young's modulus from laboratory data (Figure 18 (b)) as well as the empirical correlations are shown in Figure 18 (b). The results show that none of the static values obtained from the empirical correlations collected from the literature review were sufficiently matched with Young's modulus calibrated with the core data and only the empirical correlation which is presented by Eissa and Kazi shows a better match (indicated by the red arrow in the diagram in figure 18(b)) which has a correlation coefficient of 0.98.

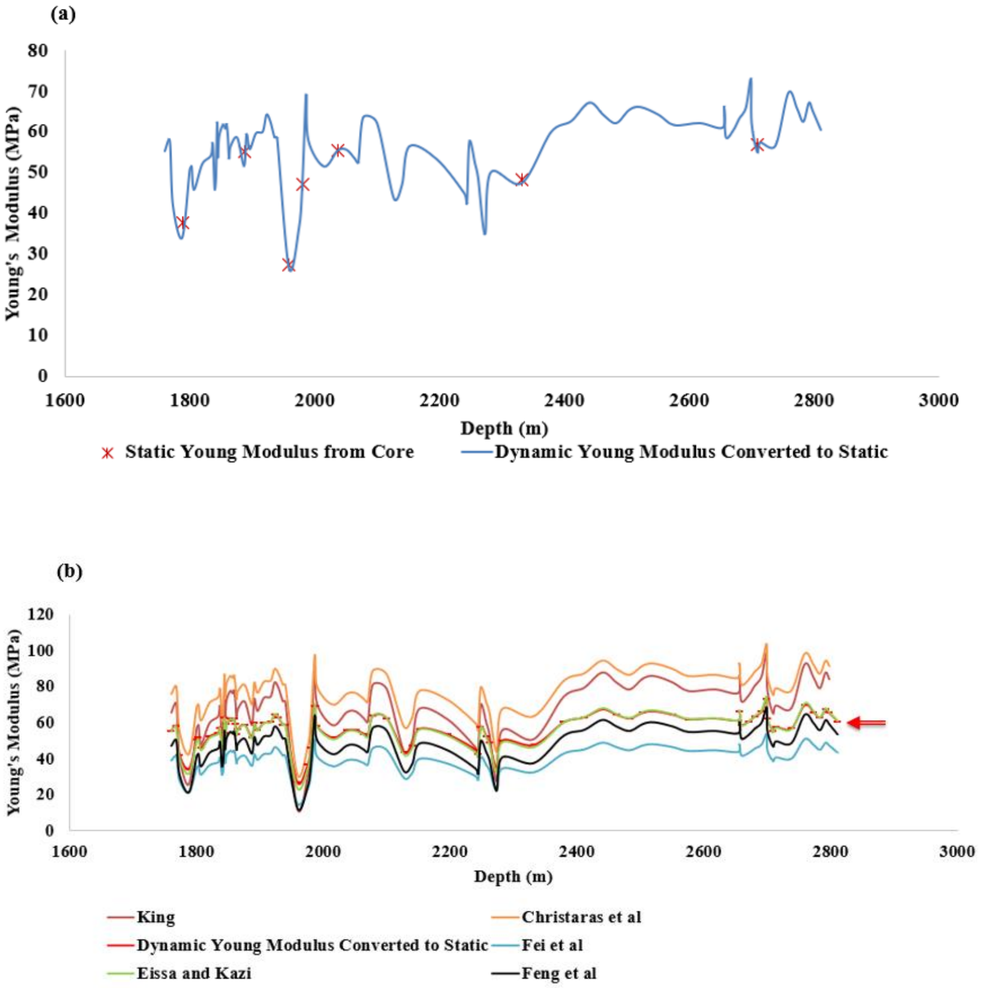

Figure 18. (a) calibrated Young's modulus by employing core data (b) comparison between static young's modulus values which is obtained from empirical correlations in previous studies with calibrated Young's modulus from core data

5. Conclusion

In this study, considering the importance of saving time, several models of machine learning methods were trained. Since the Young's modulus is of great importance in determining the field stress, this study dealt with the estimation of the dynamic Young's modulus. Young's modulus is conventionally predicted from shear and compressive wave velocities, which is not always available. In this study, a new model for predicting Young's modulus values from daily drilling reports was developed by applying different machine learning methods. These methods are linear, ANN, SVM, LSBoost, Baysian, and RF. In addition, input data to these algorithms include depth, ROP, WOB, RPM, RHOB, Φ, PP, ET, and TV. According to the presented results, the following can be concluded from the obtained research: (1) Among different mentioned machine learning algorithms, the best prediction of Young's modulus was achieved using Baysian and LSBoost methods. The correlation coefficient of Baysian for the train and test phase were 0.79 and 0.76 respectively, and the correlation coefficient of LSBoost for the train and test phase were 0.96 and 0.63 respectively. (2) The optimized parameters with the lowest prediction error used in the algorithms were presented and the best results were reported. (3) Compared to other methods in the estimation of Young's modulus, daily drilling reports are more likely to be available without extra cost and time. Therefore, the estimation of young's modulus from daily drilling reports will be very useful and beneficial. (4) By using conventional experimental correlations and core data, the resulting values were converted to static values.

Nomenclature

Acronyms	
AI	Artificial intelligence
RF	Random forest
SVM	Support vector machine
FN	Functional networks
RBF	Radial basis function
ANN	Artificial neural network
FL	Fuzzy logic
ANFIS	Adaptive neuro-fuzzy inference system
ACE	Alternating conditional expectation
PCA	Principal component analysis
DCNN	Deep convolutional neural network
LSBoost	Least Square Boosting
MLP	Multilayer perceptron
SMO	Social Media Optimization
ISDA	Intelligent Systems Designs and Applications
DDR	Daily drilling reports
D	Depth
HKHT	Hook height
HKL	Hook loud
FPWPMP	Flow pump
SPP	Stand pipe pressure
TOR	Drilling torque
W_{OUT}	Mud Weight out
CON_{IN}	Conductivity in
TMP_{IN}	Temperature in
MW_{IN}	Mud weight in
P_p	Pore pressure
P_F	Fracture pressure
PP	pump pressure
RS	Drill string rotating speed
Q	Pump flow rate
Φ	Porosity
UCS	Unconfined compressive strength
WOB	Weight on bit
RPM	String rotary speed
ROP	Rate of penetration
RHOB	Density
MW	Mud weight
TORG	Torque
TV	Tangential velocity
DE	Differential evolution
GR	Gamma rays
DSI	Dipole Sonic Imager
OOB	Out of bag (Test dataset)
MFV	Marsh funnel viscosity
ε_{trans}	Transverse strain
ε_{axial}	Axial strain
v	Poisson ratio
v_S	Static Poisson ratio
v_{dyne}	Dynamic Poisson ratio
E	Young's modulus
E_S	Static Young's modulus
E_{dyne}	Dynamic Young's modulus
ρ	Bulk density
V_S	Shear wave velocity
V_P	Compressive wave velocity
$F_0(x)$	Initial guess function
F_m	Incremental function
ρ_m	weight of the mth model
a_m	Division variable
h (x; a)	Basic learner function
$[X_i, Y_i]^N$	Training data
x and y	Input and output variables
M	Number of iterations
L	Cost function
$\tilde{y}(x)$	Predicted values
w_n	Weight of the nth learning tree
T	Weak learning tree
ϑ	Learning rate
$P(C_i)$	Probability of occurrence C
$P\ (X)$	Probability of occurrence X
$P\ (C_i \mid X)$	Probability of occurrence C occurs if X occurs
$P\ (X \mid C_i)$	Probability of occurrence X occurs if C
b	Constant of the regression function properties
$\emptyset$	Kernel function
C	Positive real constant
$\varepsilon_i\ ,\varepsilon_i^*$	Slack variables
W	weight vector
Variables	
R^2	Coefficient of determination (correlation)
MSE	Mean squared error
RMSE	Root mean square error

References

1. Al-Zubaidi, N.S. and A.K. Al-Neeamy, *3D mechanical earth model for Zubair oilfield in southern Iraq.* Journal of Petroleum Exploration and Production Technology, 2020: p. 1-13.
2. Ranjbar, A., *3D geomechanical modeling and estimating the compaction and subsidence of Fahlian reservoir formation (X-field in SW of Iran).* Arabian Journal of Geosciences, 2017. **10**: p. 116.
3. Dahab, A., et al. *Managing wellbore instability through geomechanical modeling and wellbore stability analysis*. in *54th US Rock Mechanics/Geomechanics Symposium*. 2020. OnePetro.
4. Halafawi, D. and D. Avram, *Wellbore instability prediction and performance analysis using Poroelastic modeling.* J Oil Gas Petrochem Sci, 2019. **2**(2): p. 93-106.
5. Mohammed, H.Q., A. Abbas, and H.H. Dahm. *Wellbore instability analysis for Nahr Umr Formation in southern Iraq*. in *52nd US rock mechanics/geomechanics symposium*. 2018. OnePetro.
6. Afsari, M., et al. *Using drilling and logging data for developing 1d mechanical earth model for a mature oil field to predict and mitigate wellbore stability challenges*. in *International Oil and Gas Conference and Exhibition in China*. 2010. OnePetro.
7. Charlez, P.A., A. Onaisi, and J. Lecourtier, *Wellbore stability: One of the most important engineering challenges when drilling smart wells.* Interactive drilling for fast track oilfield development: Editions Technip, 2001: p. 77-102.
8. Ma, T., et al., *Uncertainty evaluation of safe mud weight window utilizing the reliability assessment method.* Energies, 2019. **12**(5): p. 942.
9. Darvishpour, A., et al., *Wellbore stability analysis to determine the safe mud weight window for sandstone layers.* Petroleum Exploration and Development, 2019. **46**(5): p. 1031-1038.
10. Cherian, B.V., et al. *Evaluating horizontal well completion effectiveness in a field development program*. in *SPE Hydraulic Fracturing Technology Conference*. 2012. OnePetro.
11. Higgins, S.M., et al. *Anisotropic stress models improve completion design in the Baxter Shale*. in *SPE Annual Technical Conference and Exhibition*. 2008. OnePetro.
12. Assad, J.M., *The effect of orthorhombic anisotropy and its implication for oil recovery and reservoir exploitation.* Geophysical prospecting, 2005. **53**(1): p. 121-129.
13. Azin, R., A. Izadpanahi, and A. Ranjbar, *Challenges of Gas Injection*, in *Fundamentals and Practical Aspects of Gas Injection*. 2022, Springer. p. 397-413.
14. Raghavan, R. and L.Y. Chin. *Productivity changes in reservoirs with stress-dependent permeability*. in *SPE*

annual technical conference and exhibition. 2002. OnePetro.

15. Ali, A.H.A., *Watching rocks change-mechanical earth modeling.* Oilfield Review, summer, 2003. **15**: p. 22-39.
16. Fatehiboroujeni, S., A. Gopinath, and S. Goyal, *Three-dimensional nonlinear dynamics of prestressed active filaments: flapping, swirling, and flipping.* Physical Review E, 2021. **103**(1): p. 013005.
17. Atanackovic, T.M. and A. Guran, *Hooke's law*, in *Theory of elasticity for scientists and engineers*. 2000, Springer. p. 85-111.
18. Fei, W., et al., *Correlation of dynamic and static elastic parameters of rock.* Electronic Journal of Geotechnical Engineering, 2016. **21**(04): p. 1551-1560.
19. Vishkai, M., et al., *Modeling geomechanical properties in the montney formation, Alberta, Canada.* International Journal of Rock Mechanics and Mining Sciences, 2017. **96**: p. 94-105.
20. Eissa, E. and A. Kazi, *Relation between static and dynamic Young's moduli of rocks.* International Journal of Rock Mechanics and Mining & Geomechanics Abstracts, 1988. **25**(6).
21. Peng, L., L. Xinrong, and Z. Zuliang, *Mechanical property experiment and damage statistical constitutive model of Hongze rock salt in China.* Electron. J. Geotech. Eng, 2015. **20**: p. 81-94.
22. Zisman, W., *Comparison of the statically and seismologically determined elastic constants of rocks.* Proceedings of the National Academy of Sciences of the United States of America, 1933. **19**(7): p. 680.
23. Fjar, E., et al., *Petroleum related rock mechanics*. 2008: Elsevier.
24. Martínez-Martínez, J., D. Benavente, and M. García-del-Cura, *Comparison of the static and dynamic elastic modulus in carbonate rocks.* Bulletin of engineering geology and the environment, 2012. **71**(2): p. 263-268.
25. Elkatatny, S., *Application of artificial intelligence techniques to estimate the static Poisson's ratio based on wireline log data.* Journal of Energy Resources Technology, 2018. **140**(7).
26. Tariq, Z., et al. *A holistic approach to develop new rigorous empirical correlation for static Young's modulus*. in *Abu Dhabi International Petroleum Exhibition & Conference*. 2016. OnePetro.
27. Siddig, O.M., et al., *Drilling Data-Based Approach to Build a Continuous Static Elastic Moduli Profile Utilizing Artificial Intelligence Techniques.* Journal of Energy Resources Technology, 2021. **144**(2): p. 023001.
28. Siddig, O., et al., *Real-time prediction of Poisson's ratio from drilling parameters using machine learning tools.* Scientific Reports, 2021. **11**(1): p. 1-13.
29. Elkatatny , S., *Real-Time Prediction of the Dynamic Young's Modulus from the Drilling Parameters Using the Artificial Neural Networks.* Arabian Journal for Science and Engineering, 2021: p. 1-10.
30. King, M.S., *Static and dynamic elastic properties of igneous and metamorphic rocks from the Canadian shield.* 1983.
31. Najibi, A.R., et al., *Empirical relations between strength and static and dynamic elastic properties of Asmari and Sarvak limestones, two main oil reservoirs in Iran.* Journal of Petroleum Science and Engineering, 2015. **126**: p. 78-82.
32. b Heerden, W. *General relations between static and dynamic moduli of rocks*. in *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*. 1987. Elsevier.
33. Horsrud, P., *Estimating Mechanical Properties of Shale From Empirical Correlations.* SPE Drilling & Completion, 2001. **16**(02): p. 68-73.
34. Christaras, B., F. Auger, and E. Mosse, *Determination of the moduli of elasticity of rocks. Comparison of the ultrasonic velocity and mechanical resonance frequency methods with direct static methods.* Materials and Structures, 1994. **27**(4): p. 222-228.
35. Feng, C., et al., *A new empirical method based on piecewise linear model to predict static Poisson's ratio via well logs.* Journal of Petroleum Science and Engineering, 2019. **175**: p. 1-8.
36. Sozontov, A., M. Ivanova, and A. Gibadullin. *Implementation of artificial intelligence in the electric power industry*. in *E3S Web of Conferences*. 2019. EDP Sciences.
37. Kaur, S., et al., *Medical diagnostic systems using artificial intelligence (ai) algorithms: Principles and perspectives.* IEEE Access, 2020. **8**: p. 228049-228069.
38. Shamshirband, S., et al., *A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues.* Journal of Biomedical Informatics, 2020: p. 103627.
39. Tran, N.L., et al., *Application of Interpretable Machine-Learning Workflows To Identify Brittle, Fracturable, and Producible Rock in Horizontal Wells Using Surface Drilling Data.* SPE Reservoir Evaluation & Engineering, 2020. **23**(04): p. 1328-1342.
40. Salehi, S., et al., *Casing collapse risk assessment and depth prediction with a neural network system approach.* Journal of Petroleum Science and Engineering, 2009. **69**(1-2): p. 156-162.
41. Tariq, Z., et al. *A new technique to develop rock strength correlation using artificial intelligence tools*. in *SPE Reservoir Characterisation and Simulation Conference and Exhibition*. 2017. OnePetro.
42. Elkatatny, S., et al., *Development of new mathematical model for compressional and shear sonic times from wireline log data using artificial intelligence neural networks (white box).* Arabian Journal for Science & Engineering (Springer Science & Business Media BV), 2018. **43**(11).
43. Ebrahimi, A., et al., *Estimation of shear wave velocity in an Iranian oil reservoir using machine learning methods.* Journal of Petroleum Science and Engineering, 2021: p. 109841.
44. Tariq, Z., et al. *A new approach to predict failure parameters of carbonate rocks using artificial intelligence tools*. in *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*. 2017. OnePetro.
45. Elkatatny, S. and M. Mahmoud, *Development of a new correlation for bubble point pressure in oil reservoirs using artificial intelligent technique.* Arabian Journal for Science and Engineering, 2018. **43**(5): p. 2491-2500.
46. Al-AbdulJabbar, A., A.A. Mahmoud, and S. Elkatatny, *Artificial neural network model for real-time prediction of the rate of penetration while horizontally drilling natural gas-bearing sandstone formations.* Arabian Journal of Geosciences, 2021. **14**(2): p. 1-14.

47. Mahmoud, A.A., et al. *Artificial neural networks model for prediction of the rate of penetration while horizontally drilling carbonate formations*. in *54th US Rock Mechanics/Geomechanics Symposium*. 2020. OnePetro.
48. Salaheldin, E., *New approach to optimize the rate of penetration using artificial neural network.* Arabian Journal for Science and Engineering, 2018. **43**(11): p. 6297-6304.
49. Elkatatny, S., et al. *Optimization of rate of penetration using artificial intelligent techniques*. in *51st US Rock Mechanics/Geomechanics Symposium*. 2017. OnePetro.
50. Mahmoud, A.A., et al., *Estimation of static young's modulus for sandstone formation using artificial neural networks.* Energies, 2019. **12**(11): p. 2125.
51. Hadi, F., A. Eckert, and F. Almahdawi. *Real-time pore pressure prediction in depleted reservoirs using regression analysis and artificial neural networks*. in *SPE Middle East Oil and Gas Show and Conference*. 2019. OnePetro.
52. Mahmoud, A.A., et al., *Evaluation of the total organic carbon (TOC) using different artificial intelligence techniques.* Sustainability, 2019. **11**(20): p. 5643.
53. Zhu, L., et al., *Forming a new small sample deep learning model to predict total organic carbon content by combining unsupervised learning with semisupervised learning.* Applied Soft Computing, 2019. **83**: p. 105596.
54. Zhu, L., et al., *A new and reliable dual model-and data-driven TOC prediction concept: A TOC logging evaluation method using multiple overlapping methods integrated with semi-supervised deep learning.* Journal of Petroleum Science and Engineering, 2020. **188**: p. 106944.
55. Mahmoud, A.A., et al. *New robust model to evaluate the total organic carbon using fuzzy logic*. in *SPE Kuwait Oil & Gas Show and Conference*. 2019. OnePetro.
56. Zhu, L., et al., *High-precision calculation of gas saturation in organic shale pores using an intelligent fusion algorithm and a multi-mineral model.* Advances in Geo-Energy Research, 2020. **4**(2): p. 135-151.
57. Mahmoud, A.A., S. Elkatatny, and A. Al-AbdulJabbar, *Application of machine learning models for real-time prediction of the formation lithology and tops from the drilling parameters.* Journal of Petroleum Science and Engineering, 2021. **203**: p. 108574.
58. Moazzeni, A. and M.A. Haffar, *Artificial intelligence for lithology identification through real-time drilling data.* Journal of Earth Science & Climatic Change, 2015. **6**(3): p. 1-4.
59. Elkatatny, S., A. Al-AbdulJabbar, and A.A. Mahmoud, *New robust model to estimate formation tops in real time using artificial neural networks (ANN).* Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description, 2019. **60**(06): p. 825-837.
60. Mahmoud, A.A., M. Elzenary, and S. Elkatatny, *New hybrid hole cleaning model for vertical and deviated wells.* Journal of Energy Resources Technology, 2020. **142**(3).
61. Abdelgawad, K., et al., *Real-time determination of rheological properties of spud drilling fluids using a hybrid artificial intelligence technique.* Journal of Energy Resources Technology, 2019. **141**(3).
62. Ahmed S, A., et al. *Prediction of pore and fracture pressures using support vector machine*. in *International Petroleum Technology Conference*. 2019. OnePetro.
63. Ahmed, A., S. Elkatatny, and A. Ali, *Fracture Pressure Prediction Using Surface Drilling Parameters by Artificial Intelligence Techniques.* Journal of Energy Resources Technology, 2021. **143**(3): p. 033201.
64. Abdelaal, A., S. Elkatatny, and A. Abdulraheem, *Data-Driven Modeling Approach for Pore Pressure Gradient Prediction while Drilling from Drilling Parameters.* ACS omega, 2021.
65. Atashnezhad, A., A.E. Cedola, and G. Hareland. *An empirical model to estimate a critical stimulation design parameter using drilling data*. in *SPE Western Regional Meeting*. 2017. OnePetro.
66. Diaz, M.B., et al., *Drilling data from an enhanced geothermal project and its pre-processing for ROP forecasting improvement.* Geothermics, 2018. **72**: p. 348-357.
67. Diaz, M.B., et al., *Predicting rate of penetration during drilling of deep geothermal well in Korea using artificial neural networks and real-time data collection.* Journal of Natural Gas Science and Engineering, 2019. **67**: p. 225-232.
68. Mensa-Wilmot, G., B. Calhoun, and V.P. Perrin. *Formation drillability-definition, quantification and contributions to bit performance evaluation*. in *SPE/IADC Middle East Drilling Technology Conference*. 1999. OnePetro.
69. Ahmed, A., S. Elkatatny, and A. Alsaihati, *Applications of Artificial Intelligence for Static Poisson's Ratio Prediction While Drilling.* Computational Intelligence and Neuroscience, 2021. **2021**.
70. Asadi, A., *Application of artificial neural networks in prediction of uniaxial compressive strength of rocks using well logs and drilling data.* Procedia Engineering, 2017. **191**: p. 279-286.
71. Amani, A. and K. Shahbazi, *Prediction of rock strength using drilling data and sonic logs.* International Journal of Computer Applications, 2013. **81**(2).
72. Kalantari, S., H. Hashemolhosseini, and A. Baghbanan, *Estimating rock strength parameters using drilling data.* International Journal of Rock Mechanics and Mining Sciences, 2018. **104**: p. 45-52.
73. Elkatatny, S., et al. *Application of artificial intelligent techniques to determine sonic time from well logs*. in *50th US rock mechanics/geomechanics symposium*. 2016. OnePetro.
74. Mahmoud, A.A., et al. *Functional neural networks-based model for prediction of the static Young's modulus for sandstone formations*. in *54th US Rock Mechanics/Geomechanics Symposium*. 2020. OnePetro.
75. Al-Anazi, B.D., et al. *Prediction of Poisson's ratio and Young's modulus for hydrocarbon reservoirs using alternating conditional expectation Algorithm*. in *SPE Middle East oil and gas show and conference*. 2011. OnePetro.
76. Gamal, H., et al., *Rock Strength Prediction in Real-Time While Drilling Employing Random Forest and Functional Network Techniques.* Journal of Energy Resources Technology, 2021. **143**(9): p. 093004.
77. Siddig, O. and S. Elkatatny, *Workflow to build a continuous static elastic moduli profile from the drilling data using artificial intelligence techniques.* Journal of Petroleum Exploration and Production Technology, 2021: p. 1-10.
78. Abdulraheem, A., et al. *Prediction of rock mechanical parameters for hydrocarbon reservoirs using different artificial intelligence techniques*. in *SPE Saudi Arabia Section Technical Symposium*. 2009. OnePetro.
79. Tariq, Z., et al., *A rigorous data-driven approach to predict*

Poisson's ratio of carbonate rocks using a functional network. Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description, 2018. **59**(06): p. 761-777.
80. Tariq, Z., et al. *Estimation of rock mechanical parameters using artificial intelligence tools*. in *51st US Rock Mechanics/Geomechanics Symposium*. 2017. OnePetro.
81. He, M., et al., *Deep convolutional neural network for fast determination of the rock strength parameters using drilling data.* International Journal of Rock Mechanics and Mining Sciences, 2019. **123**: p. 104084.
82. Sepehr, M. and J. Cosgrove, *Role of the Kazerun Fault Zone in the formation and deformation of the Zagros Fold-Thrust Belt, Iran.* Tectonics, 2005. **24**(5).
83. KhostoTehrani, K., *Iran Geology.* Tehran: Klidar, 2005.
84. Vernant, P., et al., *Present-day crustal deformation and plate kinematics in the Middle East constrained by GPS measurements in Iran and northern Oman.* Geophysical Journal International, 2004. **157**(1): p. 381-398.
85. Aghanabati, A., *Iran Geology.* Tehran: Geological Survey & Mineral Explorations of Iran, 2004.
86. Moteyie, *Iran Geology* Tehran: Geological Survey of the country, 1374.
87. Maleki, Z., M. Arian, and A. Solgi, *Folding pattern in the Fars province, Zagros folded belt: case study on the Karbasi and Khaftar anticlines, interior Fars, Iran.* Solid Earth Discussions, 2015.
88. Alsharhan, A., *Albian Clastics in the Western Arabian Gulf Region: a Sedimentological and Petroleum-Geological Interpretation.* Journal of Petroleum Geology, 1994. **17**(3): p. 279-300.
89. Motiei, H., *Petroleum geology of the Persian Gulf.* Tehran University Press, 2009.
90. Sharma, V., S. Rai, and A. Dev, *A comprehensive study of artificial neural networks.* International Journal of Advanced research in computer science and software engineering, 2012. **2**(10).
91. Arber, S., et al., *MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure.* Cell, 1997. **88**(3): p. 393-403.
92. Bello, O., et al., *Application of artificial intelligence methods in drilling system design and operations: a review of the state of the art.* Journal of Artificial Intelligence and Soft Computing Research, 2015. **5**.
93. Chen, T., H. Chen, and R.-w. Liu, *A constructive proof and an extension of Cybenko's approximation theorem*, in *Computing science and statistics*. 1992, Springer. p. 163-168.
94. Cybenko, G., *Approximation by superpositions of a sigmoidal function.* Mathematics of control, signals and systems, 1989. **2**(4): p. 303-314.
95. Sindhwani, V., et al., *Feature selection in MLPs and SVMs based on maximum output information.* IEEE transactions on neural networks, 2004. **15**(4): p. 937-948.
96. Collobert, R. and S. Bengio. *Links between perceptrons, MLPs and SVMs*. in *Proceedings of the twenty-first international conference on Machine learning*. 2004.
97. Saljooghi, B.S. and A. Hezarkhani, *A new approach to improve permeability prediction of petroleum reservoirs using neural network adaptive wavelet (wavenet).* Journal of Petroleum Science and Engineering, 2015. **133**: p. 851-861.
98. Tariq, Z., et al. *A new artificial intelligence based empirical correlation to predict sonic travel time*. in *International Petroleum Technology Conference*. 2016. OnePetro.
99. Abdulaziz, A.M., H.A. Mahdi, and M.H. Sayyouh, *Prediction of reservoir quality using well logs and seismic attributes analysis with an artificial neural network: a case study from Farrud Reservoir, Al-Ghani Field, Libya.* Journal of Applied Geophysics, 2019. **161**: p. 239-254.
100. Bühlmann, P. and T. Hothorn, *Boosting algorithms: Regularization, prediction and model fitting.* Statistical science, 2007. **22**(4): p. 477-505.
101. Friedman, J.H., *Greedy function approximation: a gradient boosting machine.* Annals of statistics, 2001: p. 1189-1232.
102. Pearl, J., *Probabilistic reasoning in intelligent systems: networks of plausible inference*. 2014: Elsevier.
103. Uusitalo, L., *Advantages and challenges of Bayesian networks in environmental modelling.* Ecological modelling, 2007. **203**(3-4): p. 312-318.
104. Aguilera, P.A., et al., *Bayesian networks in environmental modelling.* Environmental Modelling & Software, 2011. **26**(12): p. 1376-1388.
105. Vapnik, V., *The nature of statistical learning theory*. 2013: Springer science & business media.
106. Misra, D., et al., *Application and analysis of support vector machine based simulation for runoff and sediment yield.* Biosystems engineering, 2009. **103**(4): p. 527-535.
107. Hamel, L.H., *Knowledge discovery with support vector machines*. Vol. 3. 2011: John Wiley & Sons.
108. Yu, H. and S. Kim, *SVM Tutorial-Classification, Regression and Ranking.* Handbook of Natural computing, 2012. **1**: p. 479-506.
109. Biau, G. and E. Scornet, *A random forest guided tour.* Test, 2016. **25**(2): p. 197-227.
110. Rodriguez-Galiano, V.F., et al., *An assessment of the effectiveness of a random forest classifier for land-cover classification.* ISPRS Journal of Photogrammetry and Remote Sensing, 2012. **67**: p. 93-104.
111. Ghimatgar, H., et al., *An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov model.* Journal of neuroscience methods, 2019. **324**: p. 108320.
112. Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.
113. Zou, C., et al. *A random forest regressor based uncertainty quantification of porosity estimation from multiple seismic attributes*. in *First International Meeting for Applied Geoscience & Energy*. 2021. Society of Exploration Geophysicists.
114. Ho, T.K. *Random decision forests*. in *Proceedings of 3rd international conference on document analysis and recognition*. 1995. IEEE.
115. Rodriguez-Galiano, V.F., et al., *Modelling interannual variation in the spring and autumn land surface phenology of the European forest.* Biogeosciences, 2016. **13**(11): p. 3305-3317.

مقاله تحقیقاتی

تخمین مدول یانگ با استفاده از روش‌های یادگیری ماشین و گزارش‌های روزانه حفاری

پریرخ ابراهیمی[۱،۲]، علی رنجبر[۱،۲،*]، فاطمه محمدی نیا[۱،۲]، حجت قیمتگر[۳]، عباس هاشمی‌زاده[۴]

۱. دانشکده مهندسی نفت، گاز و پتروشیمی، دانشگاه خلیج فارس، بوشهر، ایران

۲. هسته تحقیقاتی بهره برداری پایدار از منابع زیرزمینی، دانشگاه خلیج فارس، بوشهر، ایران

۳. دانشکده مهندسی سیستم‌های هوشمند و علوم داده، دانشگاه خلیج فارس، بوشهر، ایران

۴. دانشکده مهندسی شیمی، دانشگاه قم، قم، ایران

مشخصات مقاله

تاریخچه مقاله:

دریافت: ۲۸ مرداد ۱۴۰۱

دریافت پس از اصلاح: ۱۹ دی ۱۴۰۱

پذیرش نهایی: ۶ بهمن ۱۴۰۱

کلمات کلیدی:

گزارش روزانه حفاری مدول

یانگ پارامترهای ژئومکانیکی

ماشین بردار پشتیبان

شبکه عصبی مصنوعی

جنگل تصادفی

حداقل مربعات تقویت شده

*عهده‌دار مکاتبات؛علی رنجبر

رایانامه:

ali.ranjbar@pgu.ac.ir

تلفن: ۰۹۱۷۳۷۷۳۰۲۶

چکیده

پیش‌بینی پارامترهای الاستیک مانند نسبت پواسون و مدول یانگ در تعیین تنش درجا و تکمیل مدل‌سازی ژئومکانیکی از اهمیت بالایی برخوردار است. با این حال، چنین داده هایی ممکن است در بانک اطلاعات میدان نفتی موجود نباشد. بنابراین گزارش روزانه حفاری (RDD) را می‌توان به عنوان جایگزین مناسبی برای پیش بینی مدول الاستیک سنگ معرفی کرد. در این مطالعه، برای اولین بار، به تخمین مدول یانگ دینامیک با استفاده از داده‌های RDD با استفاده از انواع روش‌های یادگیری ماشین مرسوم پرداخته شده است. در این راستا از روش‌هایی نظیر ماشین بردار پشتیبان، شبکه عصبی مصنوعی، جنگل تصادفی، حداقل مربعات تقویت شده و بیزین استفاده شده است. داده‌های ورودی به این الگوریتم‌ها شامل عمق، سرعت چرخش رشته (MPR)، نرخ نفوذ (POR)، وزن روی مته (BOW)، چگالی (BOHR)، تخلخل (Φ)، فشار پمپ (PP) و سرعت مماسی (VT) است. هر یک از این الگوریتم ها از نظر دقت با استفاده از معیارهای آماری مقایسه شدند. نتایج نشان داد که با استفاده از گزارش‌های حفاری روزانه، می‌توان به برآورد خوبی از پارامترهای الاستیک دست یافت. روش‌های naisyaB و tsooBSL نسبت به سایر روش‌ها دقت عددی بالاتر و بهتری را نشان دادند..

نحوه استناد به این مقاله:

Ebrahimi P, Ranjbar A, Mohammadinia F, Ghimatgar H, Hashemizadeh A., Young's Modulus Estimation Using Machine Learning Methods and Daily Drilling Reports, Journal of Oil, Gas and Petrochemical Technology, 2023; 10(1): 1-24. DOI:10.22034/jogpt.2023.356733.1107