New Empirical Correlations to Estimate the Least Principal Stresses Using Conventional Logging Data

Ahmed Gowida, Ahmed Farid Ibrahim,* Salaheldin Elkatatny,* and Abdulwahab Ali

ABSTRACT: The maximum \( (S_{\text{max}}) \) and minimum \( (S_{\text{min}}) \) horizontal stresses are essential parameters for the well planning and hydraulic fracturing design. These stresses can be accurately measured using field tests such as the leak-off test, step-rate test, and so forth, or approximated using physics-based equations. These equations require measuring some \textit{in situ} geomechanical parameters such as the static Poisson ratio and static elastic modulus via experimental tests on retrieved core samples. However, such measurements are not usually accessible for all drilled wells. In addition, the recently proposed machine learning (ML) models are based on expensive and destructive tests. Therefore, this study aims at developing a new approach to predict the least principal stresses in a time- and cost-effective way. New models have been developed using ML approaches, that is, artificial neural network (ANN) and support vector machine (SVM), to predict \( S_{\text{min}} \) and \( S_{\text{max}} \) gradients (outputs) from well-log data (inputs). A wide-ranged set of actual field data were collected and extensively analyzed before being fed to the algorithms to train the models. The developed ANN-based models outperformed the SVM-based ones with a mean absolute average error (MAPE) not exceeding 0.30% between the actual and predicted output values. Besides, new equations have been developed to mimic the processing of the optimized networks. The new empirical equations were verified against unseen data set, resulting in a remarkably matched actual stress-gradient values, confirmed by a prediction accuracy exceeding 90% in addition to an MAPE of 0.43%. The results’ statistics confirmed the robustness of the developed equations to predict the \( S_{\text{min}} \) and \( S_{\text{max}} \) gradients with a high degree of accuracy whenever the logging data are available.

1. INTRODUCTION

The downhole formation stresses are key factors in different operations in the petroleum industry. How the stresses are concentrated in the vicinity of the wellbore directly affect the drilling operation since it controls the wellbore integrity and hence may cause many drilling-related incidents, that is, stuck bottom-hole-assembly, pack-off, and lost circulation. The availability of formation stress data that describe the wellbore stress-state condition would contribute to providing viable solutions to many integrity-related wellbore problems that may be encountered during drilling. These solutions include determining the optimum mud weight, defining the safe drilling window, specifying stable trajectories, determining casing setting depths, and so forth. Furthermore, defining the downhole stress condition or distribution is considered the cornerstone for developing a representative geomechanical model of subsurface formations whereby a broad suite of problems along different stages of the reservoir life could be addressed and resolved.

With a simplifying assumption, three mutual orthogonal principal stress components can represent the downhole stress state, that is, the overburden stress \( (S_o) \) and the least principal stresses: the maximum \( (S_{\text{max}}) \) and minimum \( (S_{\text{min}}) \) horizontal stresses. Since it is due to the compressive stress caused by the overburden formations, the vertical stress \( (S_v) \) can be estimated from the overburden formation-bulk-density log.

There are two types of techniques, that is, direct and indirect methods, to determine the least principal stresses. The direct method comprises the direct measurement of the stress state by conducting \textit{in situ} field tests such as the leak-off test, mini-fract test, step-rate test, and so forth. \( S_{\text{max}} \) cannot be directly measured using these methods, hence, theoretical (empirical) correlations are developed to estimate \( S_{\text{max}} \) depending on the values of \( S_s \) and \( S_{\text{min}} \). The main challenges of this method are being time-consuming, expensive, and usually unavailable for most of the wells. Making the matters even more challenging is that such tests are typically applied at specific depths which means there is no

Received: November 22, 2021
Accepted: April 7, 2022
Published: April 18, 2022
continuous profile of these stresses that would be available based solely on these direct tests.

On the other hand, the indirect methods involve the determination of the least principal stresses using the well-log data. Different physics-informed theoretical models, that is, uniaxial strain theory and poroelastic strain models, were developed to determine the downhole formation stresses.\textsuperscript{11-17} These models are based on lab measurements of some in situ geomechanical parameters, that is, static elastic moduli, strains, and static Poisson’s ratio. These measurements can be accurately measured from the lab tests (e.g., triaxial tests) conducted on retrieved cores that have been sampled from the downhole formations.\textsuperscript{1}

Thereafter, the measured parameters would be presented in a continuous profile form after correlating them to the conventional logging data. Besides, there is still a need for at least one direct field test, that is, the leak-off test, to incorporate the effect of tectonic stresses on the generated profiles.\textsuperscript{16-20} However, one main drawback of this technique is the high cost of retrieving such core samples to be subjected to such lab measurements. This, in turn, limits the accessibility of this kind of information for most of the drilled wells. Some recent studies introduced the application of machine learning (ML) to estimate the downhole principal stresses using the breakout data.\textsuperscript{21,22} The breakout geometries can be derived from the analysis of image logs.\textsuperscript{23} However, borehole breakouts are considered destructive techniques that are based on failure models.\textsuperscript{24} Besides, most drilled wells lack such data due to the high cost and time consumption of running these special logs. Accordingly, based on the literature, direct nondestructive techniques to determine the formation stresses are yet to be researched. Another approach was introduced by Al’Tammar and Alruwaili.\textsuperscript{25} to estimate Sh\textsubscript{min} and Sh\textsubscript{max} based on the caliper log data; however, a certainty analysis has to be incorporated into the model for geomechanical properties that are not readily available.

Therefore, a project was initiated to investigate the feasibility of ML to estimate the formation stresses using the available and easy-to-get data such as mechanical data and logging data. The results of the first phase demonstrated the ability of ML-based models to predict the in situ stresses using the mechanical drilling data.\textsuperscript{26} The second phase of the project, which is the subject of this paper, investigated the application of ML to predict the least principal stresses using logging data in a white-box version.

Therefore, this study aims at developing a new, robust tool that can estimate the gradients of the least principal stresses, Sh\textsubscript{min} and Sh\textsubscript{max} from the conventional well-log data by deploying ML approaches: artificial neural network (ANN) and support vector machine (SVM). The ML approaches have been selected due to the recent high computational capabilities of computers and the outstanding performance of such approaches to mimic and solve highly complex problems. Recently, different ML approaches have been successfully applied in the field of petroleum-related geomechanics such as predicting unconfined compressive strength,\textsuperscript{27,28} elastic parameters,\textsuperscript{29,30} and wellbore failures.\textsuperscript{31}

The novelty of this study was extended to develop state-of-the-art equations to estimate Sh\textsubscript{min} and Sh\textsubscript{max} directly from the logging data. These equations, with a detailed procedure for application, introduce the developed ML models in a white-box version to allow the reproducibility of the results, unlike the usual black-box nature of the ML models.

2. DATA ANALYSIS

This section describes the data set used for this study with summarized insights on the data preprocessing applied before proceeding with the model development.

2.1. Data Description. Field measurements, 2385 data points, were collected from two wells in a Middle East field representing a complex carbonate reservoir. These data include well-logging records and in situ maximum and minimum horizontal stresses, Sh\textsubscript{max} and Sh\textsubscript{min}. The logging data comprise gamma-ray (GR) log, formation bulk density (RHOB) log, compressional (DTC) and shear (DTS) wave transit-time log, neutron porosity (Phi) log, dynamic Poisson’s ratio (PR\textsubscript{d}), and dynamic elastic modulus (E\textsubscript{d}). The data collected from well-A have been used for training and testing the models, while the data gathered from well-B were directed to validate the developed models and verify their performance.

2.2. Data Acquisition. The stress magnitude can be estimated either by employing field tests or using developed theoretical-based equations. The equations developed based on the poroelastic model are considered the most common and applicable method to estimate the stress profile at the desired depth of the drilled wells.\textsuperscript{6,11,32} Blanton and Oslon\textsuperscript{16} were the first to introduce anisotropy in the in situ horizontal stress equation for different lithologies. Their model considers the effect of the tectonic stresses by introducing the tectonic strains into the equation. Accordingly, the least principal stresses can be estimated using eqs 1 and 2.\textsuperscript{16,32}

\[
Sh_{\text{min}} = \frac{PR_{s} - \alpha \rho \varepsilon_x}{1 - PR_{s}} + \frac{E_{d}}{1 - PR_{s}^{2}} \varepsilon \quad (1)
\]

\[
Sh_{\text{max}} = \frac{PR_{s} - \alpha \rho \varepsilon_x}{1 - PR_{s}} + \frac{E_{d}}{1 - PR_{s}^{2}} \varepsilon \quad (2)
\]

where PR\textsubscript{s} is the static Poisson’s ratio, S\textsubscript{i} is the vertical stress component, \alpha is Biot’s elastic coefficient, and \varepsilon\textsubscript{x} and \varepsilon\textsubscript{y} are the elastic strains in the Sh\textsubscript{min} and Sh\textsubscript{max} directions, respectively.

First, the vertical stress S\textsubscript{v} was estimated from the RHOB log by integrating the formation density from the surface to the depth of interest using eq 3.

\[
S_v = \int_{0}^{z} \rho(z) g \, dz \quad (3)
\]

where \rho(z) is the formation density at a certain depth of z and g is the gravitational acceleration.

Then, dynamic Poisson’s ratio (PR\textsubscript{d}) and dynamic elastic modulus (E\textsubscript{d}) were estimated based on the acoustic and RHOB logs using the formulas listed in Appendix A. The calculated E\textsubscript{d} and PR\textsubscript{d} were then correlated with E\textsubscript{i} and PR\textsubscript{s} obtained from the experimental tests conducted on the samples cored from the downhole formations.

To determine the elastic strains’ values \varepsilon\textsubscript{x} and \varepsilon\textsubscript{y}, an equal-strain assumption was initially considered for both directions before estimating the Sh\textsubscript{min} using eq 1. A field test was then used to calibrate Sh\textsubscript{min} for tectonic effects. In the case of not achieving an accurate match with the measured value, Sh\textsubscript{min} was recalculated using different ratio values (\varepsilon\textsubscript{x}/\varepsilon\textsubscript{y}). This step was repeated iteratively until an acceptable converge on Sh\textsubscript{min}
values and accurate match were achieved. Finally, the Shmin and Shmax profiles were estimated and considered as the outputs for the proposed ML models.

2.3. Statistical Descriptive Analysis. The obtained data in this study were statistically analyzed and described by deploying different statistical measures, as listed in Table 1. This helps provide better understanding of the data and its distribution. The descriptive measures listed indicated that both the logging and stress data cover a wide range with a representative distribution and hence give more confidence to capture the nature of the problem. The data ranges can be summarized as following: GR: 3.34–90.49 API unit, DTC: 44.82–66.12 μs/s, DTS: 81.28–132.47 μs/s, RHOB: 2.32–3.04 g/cm³, Phi: 0.28–0.95 fraction, Efi: 5.70–14.79 Mpsi, PRd: 0.28–0.33 fraction, Shmin: 11 292.34–32 161.71 psi, and Shmax: 12 308.02–14 599.00 psi.

2.4. Data Preprocessing. Data preprocessing is an essential step for developing ML-based models since the quality of data directly has a considerable impact on the ability of the model to learn and give accurate predictions. Therefore, the obtained data were initially preprocessed before being fed to the proposed models. The data set was first cleaned from any missing data, redundant or duplicated information, and contextual errors such as negative values and unreasonable values that do not make sense from the engineering point of view. Then, a MATLAB code was specially designed to detect and eliminate the outliers using several techniques, that is, quartiles, and so forth.

2.5. Dimensionality Reduction. It refers to the process of reducing the dimensionality of the input features, that is, the logging data, to obtain a set of principal features. Accordingly, the redundant and irrelevant input information was identified by studying the collinearity between the input features and then removed. First, the input data were normalized between 0 and 1 for better representation. Then, the correlation coefficient (R-value) was calculated between the inputs to evaluate how strongly each input linearly correlates with the others (Table 2). In the case of having two or more features that have an R-value of more than 0.95, only one of them would be considered, and the others would be excluded. Therefore, only GR, RHOB, and DTC were selected to feed the proposed models after excluding the others that have almost the same distribution as the selected ones (Figure 1).

Moreover, taking the square root (Sqrt) of the GR values, square-root transformation reduced its skewness from 0.63 to −0.07. It approached zero, which is indicative of being more like a normal distribution. Therefore, Sqrt(GR) values were considered instead of GR values as an input feature.

2.6. Correlation Analytics. Pearson’s correlation coefficient was used to investigate the relative importance between each input feature and the outputs. This correlation helps identify to what extent the output is dependent on each input feature. The R-value between Shmin and the selected inputs did not exceed 0.29. As an attempt to enhance this value, the formation depth was integrated into the stress profile to express it as a stress gradient profile instead. Studying the correlation between each feature and the Shmin gradient, Figure 2a shows a significant increase in the R-value from 0.29, −0.27, and 0.12 to −0.53, 0.60, and 0.21 for Sqrt(GR), DTC, and RHOB, respectively, compared to the initial case with the Shmin values.

Similarly, the Shmax gradient was found to have a relatively higher R-value with the input features compared to Shmax R-values, as shown in Figure 2b. Therefore, the Shmin and Shmax gradients were considered the proposed models’ outputs instead of the absolute Shmin and Shmax values. The formula used to calculate Pearson’s correlation coefficient is presented in Appendix A.

3. MODEL DEVELOPMENT

The proposed models were then developed using the preprocessed data set by employing ANN and SVM techniques to predict the Shmin and Shmax gradients based on the selected conventional logging data; GR, DTC, and RHOB.

3.1. Artificial Neural Network. ANN as a supervised-learning technique is recently well-known for its high capability of modeling several engineering problems with a high degree of complexity. The basic architecture of a neural network typically consists of three types of layers: the input layer, hidden layer(s), and output layer. The input features are assigned to the input layer that has weighted connections with the hidden layer(s). The neurons in the hidden layer process the input data before being transferred through the network connections to the output layer to ultimately produce the output in the output layer. The optimization process of the network aims at tuning the weights of the network connections as well as the biases to yield the lowest possible error for a given network configuration.

3.2. Support Vector Machine. SVM is one of the most common ML techniques, well-known for its high capability to deal with classification and regression applications with a high degree of complexity. It follows the supervised learning approach while carrying out the transformation of the input data set into a higher-degree dimensional (n-dimensional) feature space whereby more space would be available for...
training instances to achieve the optimal hyper-plane. Recent studies have employed the SVM technique in estimating several petroleum-related parameters and in geomechanics-related applications. Recently, many studies employed the SVM technique in estimating several petroleum-related parameters and in geomechanics-related applications.45–49

4. RESULTS AND DISCUSSION

4.1. ANN-Based Model Development. In this study, ANN was employed to develop new models that can estimate the $Sh_{min}$ and $Sh_{max}$ gradients based on the well-log data as feeding inputs. The obtained data set was initially divided into three main categories: training, validation, and testing sets. Typically, multiple models are trained using the training set with different hyper-parameters before being tested internally utilizing the validation set to evaluate the selected hyper-parameters. The developed model with those hyper-parameters that yield acceptable prediction accuracy on the validation set is then tested using the testing set to evaluate the generalization error of the trained model.39

Ratios ranging from 70 to 90% were tested for the training set, and for each trial, the rest of the data was split using a one-to-one ratio for the testing and validation sets. Meanwhile, different combinations of the ANN parameters were tested to optimize the model. Table 3 lists the ANN-parameter options that have been tested in addition to the selected (optimized) ones.

The gradient descent algorithm was implemented while iteratively updating the network parameters in the gradient direction of the objective function. The process includes

---

**Figure 1.** Graphic display of the distribution of the normalized logging data where the $y$-axis represents the normalized data values and the $x$-axis represents the data index.

| GR | Phi | $E_d$ |
|----|-----|-------|
| DTC | DTS | PR$_d$ |

| RHOB | $Sh_{max}$ and $Sh_{min}$ |
|------|---------------------------|

Index
considering random values of the model hyperparameters and iteratively adjusting them using the available options to eventually reduce the loss function over a series of trials (epochs). The model’s hyper-parameters are updated through each iteration to minimize the loss of the next iteration using the back propagation technique.

A MATLAB code was developed to test different scenarios while optimizing the network. Each scenario includes different combinations of the available options of the ANN parameters. The prediction for each case was evaluated in terms of the R-value to assess the collinearity between the predicted and actual output values. In addition, the prediction error was evaluated using the mean absolute percentage error (MAPE) and root-mean-squared error (RMSE) between the predicted and observed output values for the training, validation, and testing processes. Achieving the highest R-value besides the lowest MAPE and RMSE was the objective criteria to select the optimized parameters of the network. The mathematical formulas used to calculate MAPE and RMSE are stated in Appendix A.

4.1 Shmin Gradient Prediction. The tuning process of the developed Shmin gradient model results in a network architecture of three layers: one input layer including the input features [Sqrt(GR); DTC, and RHOB], one hidden layer with 30 neurons, and one output (Shmin gradient) layer. The developed model was trained by the Levenberg Marquardt algorithm (trainlm) with a learning rate of 0.05 using a transfer function of tan-sigmoidal type for the input layer and a linear function for the output layer. Figure 3 shows a typical architecture schematic of the developed ANN-based models. The crossplots between the predicted and actual Shmin gradients, Figure 4, showed a significant match with an R-value of 0.90 and MAPE not exceeding 0.14% both for the training and testing processes.

After fitting a regression model, the prediction residuals have been checked to ensure reliable regression results. Therefore, the residuals of the Shmin-gradient model were plotted versus the fitted values, as depicted in Figure 5a, which shows the random scattering of the residuals around zero. The residual histograms were also found to be more-like normally distributed (Figure 5b), which demonstrates that all the fitted values have almost the same degree of scattering.

4.1.2. Shmax Gradient Prediction. Similarly, the optimized network for predicting the Shmax gradient contained one hidden layer with 15 neurons. The model was trained with a learning rate of 0.15 using trainlm as a learning algorithm.

Figure 3. Typical architecture of the developed ANN-based models.

Table 3. Tested Options for Optimizing the Developed ANN-Based Models

| parameter                  | tested options/ranges | Shmin gradient model | Shmax gradient model |
|---------------------------|-----------------------|----------------------|----------------------|
| number of hidden layers   | 1–4                   | single hidden layer  |
| number of neurons in each layer | 5–40          | 30                   | 15                   |
| split ratio               | 70%–90% (for training set) the rest was divided by 1-to-1 ratio for validation and testing | 0.8/0.1/0.1          |
| training algorithms       | trainlm, trainbfg, trainrp | trainlm              |
| transfer function         | tansig, logsig, elliotsg | tansig               |
| learning rate             | 0.01–0.9              | 0.05                 | 0.15                 |
Figure 4. Crossplots between the actual and predicted $\text{Sh}_{\text{min}}$ gradients for the developed ANN-based model for (a) training and (b) testing processes.

Figure 5. Analysis of the prediction residuals of the $\text{Sh}_{\text{min}}$-gradient ANN-based model: (a) residuals vs fitted values and (b) histogram of the prediction residuals.

Figure 6. Crossplots between the actual and predicted $\text{Sh}_{\text{max}}$ gradients for the developed ANN-based model for (a) training and (b) testing processes.
The narrow scatter of the points along the 45-line in the crossplots is shown in Figure 6, indicating the agreement between the observed Shmax gradient and the predicted ones for both the training and testing. This is further verified by the low MAPE 0.30% between the observed and predicted values for the testing process. In addition, the average R-value is 0.98 for both. The evaluation metrics (R-value, MAPE, and RMSE) listed in Table 4 describe the accuracy of the ANN-based models. Furthermore, plotting the model prediction residuals versus the fitted values showed a scattered pattern around zero, Figure 7a, with approximately a normal distribution in the residual histogram plot depicted in Figure 7b. These measures indicate the stable prediction (regression) performance of the developed model.

### 4.2. SVM-Based Model Development.

The same data set was used for building the SVM-based models to estimate the Shmin and Shmax gradients using the same input features. For optimizing the SVM-based models, both Gaussian and polynomial kernel functions were tested with different SVM-model optimizing parameters; epsilon, lambda, kernel option, C-parameter, and verbose. The model was trained using 70% of the obtained data, while the rest were used for the validation and testing processes with a one-to-one ratio. For both the Shmin- and Shmax-gradient models, the sensitivity analysis shows that the epsilon, lambda, and verbose parameters did not significantly impact prediction accuracy. The Gaussian kernel function yielded better prediction performance regarding the R-value between the predicted and actual output values than the polynomial function. Varying kernel options from one to nine showed that a kernel option of 3.5 gave the best prediction performance with the lowest MAPE for both the Shmin- and Shmax-gradient models. A C-parameter of 400 was selected for the Shmin gradient model, while 600 was chosen for the Shmax-gradient model. Increasing the C-parameter value beyond the values chosen resulted in an over-fitting problem in the developed models indicated by low training error while, conversely, very high errors in the testing process. These selected values of the SVM-based model parameters yielded the best prediction performance during the testing process in terms of the R-value of 0.86 and 0.97 and MAPE values of 0.16 and 0.41% between the predicted and the actual values for the Shmin and Shmax gradient models, respectively. The statistical parameters (R-value, MAPE, and RMSE) describing the performance of the SVM-based models to estimate the Shmin and Shmax gradients are listed in Table 4.

Table 5 summarizes the selected SVM parameters for the developed Shmin- and Shmax-gradient models. Figures 8 and 9 show the crossplots between the predicted and observed output values for model development processes (training and testing).
Comparing the prediction performance of both ANN and SVM models in the testing data set showed that the developed ANN-based models outperformed the SVM-based ones while predicting Sh\textsubscript{min} and Sh\textsubscript{max} gradients. The developed ANN-based models yielded better predictions for the testing process of the developed models regarding higher R-values of 0.92 and 0.98 and lower MAPE values of 0.14 and 0.30% between the predicted and actual Sh\textsubscript{min} and Sh\textsubscript{max} gradients. However, the

Figure 8. Crossplots between the actual and predicted Sh\textsubscript{min} gradients for the developed SVM-based models for (a) training and (b) testing processes.

Figure 9. Crossplots between the actual and predicted Sh\textsubscript{max} gradients for the developed SVM-based models for (a) training and (b) testing processes.

Figure 10. Comparison of the prediction performance between the developed (Sh\textsubscript{min} gradient) ANN-based and SVM-based models in terms of (a) R-value and (b) MAPE for training, validation, and testing processes.
predictions of the developed SVM-based models resulted in R-values of 0.86 and 0.97 with MAPE values of 0.16 and 0.41% for the Shmin and Shmax gradients models, respectively, Figures 10 and 11. Furthermore, the ANN approach has the privilege of having the potential to extract imitating equations to the predictions of the developed SVM-based models.

Table 6. Extracted Weights and Biases to Be Used in Eq 7 for Estimating the Shmax Gradient

| i | j | Wij | b1j | b2j |
|---|---|------|------|------|
| 2 | 1 | 0.000 | 2.934 | 2.366 | 2.098 | 4.579 | -0.870 |
| 1 | 3 | -9.341 | 0.527 | 1.072 | -0.298 | -3.932 |
| 3 | 0.043 | -6.361 | -0.428 | -0.876 | -5.893 |
| 4 | -3.070 | -2.316 | -2.405 | 0.664 | 2.966 |
| 5 | 2.943 | 1.749 | -2.839 | 0.685 | -3.041 |
| 6 | -1.677 | -4.048 | 0.278 | 0.127 | 2.709 |
| 7 | -0.767 | -4.177 | -4.646 | -0.340 | 1.410 |
| 8 | -3.120 | -3.061 | 0.288 | 0.945 | 1.909 |
| 9 | 0.968 | 2.507 | 2.958 | 0.044 | -1.716 |
| 10 | -3.968 | 1.687 | 1.129 | -0.330 | 1.137 |
| 11 | -1.476 | 4.656 | 2.314 | 0.339 | 2.230 |
| 12 | 3.189 | 2.455 | -1.452 | 0.154 | -1.175 |
| 13 | -3.588 | 0.998 | 2.001 | 0.278 | 0.889 |
| 14 | -0.228 | 1.546 | 3.449 | 0.625 | 1.180 |
| 15 | -1.189 | 3.067 | -2.984 | 0.213 | 0.542 |
| 16 | 1.895 | -2.761 | 2.420 | 0.148 | 1.167 |
| 17 | 3.512 | 1.644 | 2.954 | 0.251 | 0.944 |
| 18 | -1.783 | -1.520 | -0.185 | 0.752 | -1.164 |
| 19 | -4.342 | -0.693 | -3.299 | 0.380 | -0.803 |
| 20 | -4.097 | -0.671 | -2.223 | -0.496 | -0.770 |
| 21 | 2.793 | -3.729 | 0.564 | -0.697 | 1.154 |
| 22 | 1.768 | -2.186 | 3.675 | -0.134 | 1.558 |
| 23 | -0.492 | -1.169 | 5.038 | -0.751 | -3.895 |
| 24 | -3.288 | 0.868 | -2.794 | 0.185 | -2.401 |
| 25 | 3.086 | 0.805 | 3.362 | 0.112 | 1.596 |
| 26 | 4.075 | -1.966 | 1.095 | -0.105 | 2.999 |
| 27 | -4.754 | 1.211 | -1.468 | -0.870 | -2.993 |
| 28 | -0.508 | 1.499 | -4.257 | -4.311 | -3.777 |
| 29 | 2.612 | 4.141 | 2.946 | -0.492 | 4.200 |
| 30 | 0.698 | -1.903 | 5.501 | 0.693 | -4.498 |

was the development of new empirical equations that can be used to estimate the Shmin and Shmax gradients without needing to run the MATLAB codes. Accordingly, Shmin and Shmax gradients can be calculated using the novel ANN-based eqs 4 and 5, respectively.

\[
\text{Shmin gradient} = 0.012 \times (\text{Shmin gradient})_{\text{normalized}} + 0.881 \\
\text{Shmax gradient} = 0.056 \times (\text{Shmax gradient})_{\text{normalized}} + 0.995
\]

The subscript “normalized” refers to the normalized form of the Shmin and Shmax gradients, and the input parameters should be first normalized using the point-slope form in eq 6.

\[
X_{\text{normalized}} = 2 \times \left( \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \right) - 1
\]

where X is the actual value of the input parameter, Xmin and Xmax are the minimum and maximum values of the input features, respectively, and Xnormalized is the normalized form of the input parameter. The normalized form of the Shmin and
Shmax gradients in eqs 4 and 5 can be calculated using eqs 7 and 8.

\[
\left(\text{Sh}_{\text{min}} \text{ gradient}\right)_{\text{normalized}} = \left(\sum_{i=1}^{N} w_{2,i} \left[ 1 + e^{-2(\sqrt{\text{GR}} - \text{DTC}_i) + (w_{1,1} \times \text{DTC}_i) + (w_{1,3} \times \text{RHOB}_i) + b_1} - 1 \right] + b_2 \right)^2
\]

\[
\left(\text{Sh}_{\text{max}} \text{ gradient}\right)_{\text{normalized}} = \sum_{i=1}^{N} w_{2,i} \left[ 1 + e^{-2(\sqrt{\text{GR}} - \text{DTC}_i) + (w_{1,1} \times \text{DTC}_i) + (w_{1,3} \times \text{RHOB}_i) + b_1} - 1 \right] + b_2
\]

The \([\sqrt{\text{GR}}]_{i}\), DTC\(_i\), and RHOB\(_i\) represent the normalized forms of the input parameters obtained using eq 6. These equations were established to mimic the developed ANN-based models utilizing the tuned weights and biases of the optimized networks. The weights and biases of the developed Sh\(_{\text{min}}\) and Sh\(_{\text{max}}\) models in eqs 7 and 8 are listed in Figure 12.
4.4. Model Verification. For further investigation of the performance of the developed equations, 456 (unseen) data points from well B were used to evaluate the performance of the developed equations. These data involved the logging measurements (GR, RHOB, and DTC) and the corresponding Shmin and Shmax gradients. The logging data were fed as inputs for the developed ANN-based equations, and the results were then compared with the actual stress-gradient values. The prediction results of both the Shmin and Shmax gradients remarkably matched the actual values, confirmed by MAPE values of 0.18 and 0.43%; besides, R-value exceeds 0.90 for the Shmin and Shmax predictions, respectively, Figure 12. These results demonstrate the outstanding performance of the developed ANN-based equations to develop continuous profiles of the Shmin and Shmax with high accuracy whenever the well-logging data are available.

Having continuous profiles of the least principal stresses for the drilled wells could help provide practical solutions to several borehole instability issues that may affect the well integrity. Besides, such data would help develop a comprehensive geomechanical model of the subsurface formations. As a result, a broad suite of problems along different stages of the well life could be addressed and avoided.

It should be highlighted that the application of the developed correlation is more recommended for carbonate formations from which most of the data used in developing the models were obtained. This can be explained as other formation types may have different log responses to the geomechanical properties that control the downhole stress distributions. Therefore, some errors might be expected upon the application for different formation lithologies. Moreover, it is recommended to employ the developed equations using inputs within the range and the same units listed in Table 1 to ensure reliable results.

## CONCLUSIONS

New models were developed using two ML techniques, ANN and SVM, to predict the maximum (Shmax) and minimum (Shmin) horizontal stress gradients. The developed models used the conventional logging data: GR, RHOB, and DTC as feeding inputs to the algorithms. The findings of this research can be highlighted as follows:

- The prediction performance of the developed models by ANN surpassed the SVM-based ones with accuracy exceeding 90% and a MAPE of 0.30%.
- Novel equations were established according to the tuned weights and biases of the optimized neural networks. These equations can estimate the Shmin and Shmax gradients directly from the logging data.
- The new equations were validated using a different data set achieving an obvious match between the predicted and actual stress-gradient values with MAPE not exceeding 0.43%. The results reflect the robustness of the new equations to accurately estimate the Shmin and Shmax gradients directly from the well-logging data.

## APPENDIX A

### Dynamic Elastic Modulus (Ed) Equation

\[
E_d = \frac{\text{RHOB} \times \left[ 3 \left( \frac{\text{DTS}}{\text{DTC}} \right)^2 - 4 \right]}{\left( \frac{\text{DTS}}{\text{DTC}} \right)^2 - 1}
\]  

(A1)

where Ed is the dynamic elastic modulus, RHOB is the formation bulk density, and DTC and DTS are the compressional and shear wave transit times (in μs/ft), respectively.\(^{17}\)

### Dynamic Poisson’s Ratio (PRd) Equation

\[
\text{PR}_d = \frac{\left( \frac{\text{DTS}}{\text{DTC}} \right)^2 - 2}{2 \left( \frac{\text{DTS}}{\text{DTC}} \right)^2 - 1}
\]

(A2)

where DTC and DTS are the compressional and shear waves transit times (in μs/ft), respectively.\(^{17}\)

### Pearson Correlation Coefficient (R-Value)

The formula used to calculate the Pearson correlation coefficient (R-value) between two variables (x and y) using a “k” number of the data points is

\[
R = \frac{k \sum_{i=1}^{N} xy - (\sum_{i=1}^{N} x)(\sum_{i=1}^{N} y)}{\sqrt{k(\sum_{i=1}^{N} x^2) - (\sum_{i=1}^{N} x)^2} \sqrt{k(\sum_{i=1}^{N} b^2) - (\sum_{i=1}^{N} b)^2}}
\]

(A3)

### Mean Absolute Percentage Error

\[
\text{MAPE} (%) = \frac{\sum_{i=1}^{N} \left| \frac{X_{\text{measured}} - X_{\text{predicted}}}{X_{\text{measured}}} \right| \times 100}{N}
\]

(A4)

where \(X_{\text{measured}}\) and \(X_{\text{predicted}}\) are the actual and predicted values of the parameter, respectively, and \(N\) is the number of the data points.

### Root-Mean-Squared Error

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (X_{\text{measured}} - X_{\text{predicted}})^2}{N}}
\]

(A5)

where \(X_{\text{measured}}\) and \(X_{\text{predicted}}\) are the actual and predicted values of the parameter, respectively, and \(N\) is the number of the data points.

## AUTHOR INFORMATION

### Corresponding Authors

Ahmed Farid Ibrahim — College of Petroleum Engineering & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; orcid.org/0000-0001-7258-8542; Email: ahmed.ibrahim@kfupm.edu.sa

Salaheldin Elkatatny — College of Petroleum Engineering & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; orcid.org/0000-0002-7209-3715; Email: elkatatny@kfupm.edu.sa

### Authors

Ahmed Gowida — College of Petroleum Engineering & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

Abdulwahab Ali — College of Petroleum Engineering & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia
Funding
This research received no external funding.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS
The authors would like to thank King Fahd University of Petroleum & Minerals (KFUPM) for employing its resources in conducting this work.

NOMENCLATURE

- Shmin: minimum horizontal stress
- Shmax: maximum horizontal stress
- ML: machine learning
- ANN: artificial neural network
- SVM: support vector machine
- R-value: correlation coefficient
- MAPE: mean absolute percentage error
- RMSE: root-mean-squared error
- GR: gamma ray
- RHOB: formation bulk density
- DTC: compressional wave transit time
- DTS: shear wave transit time
- PRd: dynamic Poisson’s ratio
- Ed: dynamic elastic modulus
- trainlm: Levenberg–Marquardt
- trainbfg: BFGS quasi-Newton
- trainrp: resilient backpropagation
- trainscg: scaled conjugate gradient
- traincgb: conjugate gradient with Powell/Beale restarts
- traincfg: Fletcher–Powell conjugate gradient
- traincgp: Polak–Ribiére conjugate gradient
- trainoss: one step secant
- traingdx: variable learning rate backpropagation
- tansig: hyperbolic tangent sigmoid transfer function
- logsig: log-sigmoid transfer function
- elliotsg: elliot symmetric sigmoid transfer function
- radbas: radial basis transfer function
- hardlim: symmetric hard-limit transfer function
- satlin: symmetric saturating linear transfer function
- N: total number of neurons in the ANN hidden layer
- W1: matrix of the optimized weights between the ANN input and hidden layers
- W2: matrix of the optimized weights between the ANN hidden and output layers
- b1: vector of the optimized biases between the input and hidden layers
- b2: optimized bias between the hidden and output layers
- [Sqrt(GR)]n: normalized value of Sqrt(GR)
- DTCn: normalized value of DTC
- RHOBn: normalized value of RHOB

REFERENCES

1. Al-Zankawi, O.; Belhouchet, M.; Abdessalem, A. Real-Time Integration of Geo-Mechanics to Overcome Drilling Challenges and Low NPT. SPE Kuwait Oil & Gas Show and Conference; OnePetro, 2017.
2. Zoback, M. D.; Barton, C. A.; Brudy, M.; Castillo, D. A.; Finkbeiner, T.; Grollimund, B. R.; Moos, D. B.; Peska, P.; Ward, C. D.; Wiprut, D. J. Determination of Stress Orientation and Magnitude in Deep Wells. Int. J. Rock Mech. Min. Sci. 2003, 40, 1049–1076.
3. Warminska, N. R.; Teufel, L. W. In-Situ Stresses in Low-Permeability, Nonmarine Rocks. J. Pet. Technol. 1989, 41, 405–414.
4. Bell, J. S. Petro Geoscience 2. In Situ Stresses in Sedimentary Rocks (Part 2): Applications of Stress Measurements. Geosci. Can. 1996, 23, 135.
5. Willson, S. M.; Moschovidis, Z. A.; Cameron, J. R.; Palmer, I. D. New Model for Predicting the Rate of Sand Production. SPE/ISRM Rock Mechanics Conference; Society of Petroleum Engineers, 2002.
6. Molaghab, A.; Taherynia, M. H.; Fatemi Aghda, S. M.; Fahimifar, A. Determination of Minimum and Maximum Stress Profiles Using Wellbore Failure Evidences: A Case Study—a Deep Oil Well in the Southwest of Iran. J. Pet. Explor. Prod. Technol. 2017, 7, 707–715.
7. Maleki, S.; Gholami, R.; Rasouli, V.; Moradzadeh, A.; Riabi, R. G.; Sadagzadeh, F. Comparison of Different Failure Criteria in Prediction of Safe Mud Weigh Window in Drilling Practice. Earth-Sci. Rev. 2014, 136, 36–58.
8. McGarr, A.; Gay, N. C. State of Stress in the Earth’s Crust. Annu. Rev. Earth Planet. Sci. 1978, 6, 405.
9. Nolte, K. G. Principles for Fracture Design Based on Pressure Analysis. Soc. Pet. Eng., Prod. Eng. 1988, 3, 22–30.
10. Carnegie, A.; Thomas, M.; Elfnik, M. S.; Hamawi, M.; Akbar, M.; Burton, M. An Advanced Method of Determining Insitu Reservoir Stresses: Wireline Conveyed Micro-Fracturing. Adu Dhabi Interna-
11. Zoback, M. D. Reservoir Geomechanics; Cambridge University Press, 2010.
12. Binh, N. T. T.; Tokunaga, T.; Goulty, N. R.; Son, H. P.; Van Binh, M. Stress State in the Cuu Long and Nam Con Son Basins, Offshore Vietnam. Mar. Pet. Geol. 2011, 28, 973–979.
13. Zang, A.; Stephansson, O.; Heidbach, O.; Janouschkowetz, S. World Stress Map Database as a Resource for Rock Mechanics and Rock Engineering. Geotech. Geol. Eng. 2012, 30, 625–646.
14. Terzaghi, K. Erdbaumechanik Auf Bodenphysikalischer Grund-
15. Anderson, E. M. The Dynamics of Faulting. Trans. Edinburgh Geol. Soc. 1905, 8, 387–402.
16. Blanton, T. L.; Olson, J. E. Stress Magnitudes from Logs: Effects of Tectonic Strains and Temperature. SPE Reservoir Eval. Eng. 1999, 2, 62–68.
17. Fjar, E.; Holt, R. M.; Raaen, A. M.; Horsrud, P. Petroleum Related Rock Mechanics, 2nd ed.; Elsevier, 2008; Vol. 53.
18. Meng, Z.; Zhang, J.; Wang, R. In-Situ Stress, Pore Pressure and Stress-Dependent Permeability in the Southern Qinshui Basin. J. Rock Mech. Min. Sci. 2011, 48, 122–131.
19. Rasouli, V.; Pallikathakothil, Z.; Mawuli, E. The Influence of Perturbed Stresses near Faults on Drilling Strategy: A Case Study in Blacktip Field, North Australia. J. Pet. Sci. Eng. 2011, 76, 37–50.
20. Fan, X.; Gong, M.; Zhang, Q.; Wang, J.; Bai, L.; Chen, Y. Prediction of the Horizontal Stress of the Tight Sandstone Formation in Eastern Sulige of China. J. Pet. Sci. Eng. 2014, 113, 72–80.
21. Lin, H.; Singh, S.; Oh, J.; Canbulat, I.; Kang, W. H.; Hebblewhite, B.; Stacey, T. R. A Combined Approach for Estimating Horizontal Principal Stress Magnitudes from Borehole Breakout Data via Artificial Neural Network and Rock Failure Criterion. Int. J. Rock Mech. Min. Sci. 2020, 136, 104539.
22. Lin, H.; Kang, W.-H.; Oh, J.; Canbulat, I. Estimation of In-Situ Maximum Horizontal Principal Stress Magnitudes from Borehole Breakout Data Using Machine Learning. Int. J. Rock Mech. Min. Sci. 2020, 126, 104199.
23. Tingay, M.; Reinecker, J.; Müller, B. Borehole Breakout and Drilling-Induced Fracture Analysis from Image Logs; World Stress Map Project, 2008; Vol. 1–8.
(24) Sinha, B. K.; Wang, J.; Kira, S.; Li, J.; Pstre, V.; Bratton, T.; Sanders, M.; Jun, C. Estimation of Formation Stresses Using Borehole Sonic Data. SPWLA 49th Annual Logging Symposium; OnePetro, 2008.
(25) AlTammar, M. J.; Alruwaili, K. M. Integrating Monte Carlo Simulation, Machine Learning and Physics-Based Solutions to Estimate In-Situ Stresses. ARMA/DGS/SEG International Geosciences Symposium, 2020, Vol. 15.
(26) Gowida, A.; Ibrahim, A. F.; Elkatatny, S.; Ali, A. Prediction of the Least Principal Stresses Using Drilling Data: A Machine Learning Application. Comput. Intell. Neurosci. 2021, 8865827.
(27) Gowida, A.; Elkatatny, S.; Gamal, H. Unconfined Compressive Strength (UCS) Prediction in Real-Time While Drilling Using Artificial Intelligence Tools. Neural Comput. Appl. 2021, 33, 8043−8054.
(28) Chemmakh, A. Machine Learning Predictive Models to Estimate the UCS and Tensile Strength of Rocks in Bakken Field. SPE Annual Technical Conference and Exhibition; OnePetro, 2021.
(29) Gowida, A.; Moussa, T.; Elkatatny, S.; Ali, A. A Hybrid Artificial Intelligence Model to Predict the Elastic Behavior of Sandstone Rocks. Sustainability 2019, 11, 5283.
(30) Song, L.; Liu, Z.; Li, C.; Ning, C.; Hu, Y.; Wang, Y.; Hong, F.; Tang, W.; Zhuang, Y.; Zhang, R.; Zhang, Y.; Zhang, Q. Prediction and Analysis of Geomechanical Properties of Jimusaer Shale Using a Machine Learning Approach. SPWLA 62nd Annual Logging Symposium; OnePetro, 2021.
(31) Albahrani, H. I. H. An Automated Drilling Geomechanics Simulator Using Machine-Learning Assisted Elasto-Plastic Finite Element Model, Ph.D. Thesis, Texas A&M University, 2020.
(32) Abidideh, M.; Fathabadi, M. R. Analysis of Stress Field and Determination of Safe Mud Window in Borehole Drilling (Case Study: SW Iran). J. Pet. Explor. Prod. Technol. 2013, 3, 105−110.
(33) Garcia, S.; Luengo, J.; Herrera, F. Data Preprocessing in Data Mining: Intelligent Systems Reference Library; Springer International Publishing; Cham, 2015; Vol. 72.
(34) Kotsiantis, S. B.; Kanellopoulos, D.; Pintelas, P. E. Data Preprocessing for Supervised Learning Abstract — Many Factors Affect the Success Of. Int. J. Comput. Sci. 2006, 2, 111−117.
(35) Sedgwick, P. Pearson’s Correlation Coefficient. BMJ 2012, 345, No. e4483.
(36) Lippmann, R. An Introduction to Computing with Neural Nets. IEEE ASSP Mag. 1987, 4, 4−22.
(37) Hinton, G. E.; Osindero, S.; Teh, Y.-W. A Fast Learning Algorithm for Deep Belief Nets. Neural Comput. 2006, 18, 1527−1554.
(38) Nakamoto, P. Neural Networks & Deep Learning; Createspace Independent Publishing, 2017.
(39) Xu, Y.; Goodacre, R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. J. Anal. Test. 2018, 2, 249−262.
(40) Frost, J. Introduction to Statistics: An Intuitive Guide; Statistics by Jim Publishing; State College, PA, USA, 2019; pp 196−204.
(41) Durgesh, K. S.; Lekha, B. Data Classification Using Support Vector Machine. J. Theor. Appl. Inf. Technol. 2010, 12, 1−7.
(42) Gholami, R.; Fakhari, N. Support Vector Machine: Principles, Parameters, and Applications. Handbook of Neural Computation; Elsevier, 2017; pp 515−535.
(43) Suthaharan, S. Support Vector Machine. Machine Learning Models and Algorithms for Big Data Classification; Springer, 2016; pp 207−235.
(44) Pisner, D. A.; Schnyer, D. M. Support Vector Machine. Machine Learning; Elsevier, 2020; pp 101−121.
(45) Zhao, H.-b.; Yin, S. Geomechanical Parameters Identification by Particle Swarm Optimization and Support Vector Machine. Appl. Math. Model. 2009, 33, 3997−4012.
(46) Jahanbakhshi, R.; Keshavarzi, R.; Aliyari Shoorehdeli, M.; Emamzadeh, A. Intelligent Prediction of Differential Pipe Sticking by Support Vector Machine Compared with Conventional Artificial Neural Networks: An Example of Iranian Offshore Oil Fields. SPE Drill. Complet. 2012, 27, 586−595.