Petroleum and Coal

Article Open Access

Real-Time Drillstring Vibration Prediction Using Surface Drilling Data through Machine Learning Methods

*Mostafa Bayan**, *Said K. Elsayed, Shady G. Elramah*

Petroleum Engineering Department, Faculty of Petroleum and Mining Engineering, Suez University, Suez, Egypt

Received March 6, 2025; Accepted May 28, 2025

Abstract

Drillstring vibration is a significant problem during drilling operations that can pose severe consequences such as downhole tool failure, borehole instability, loss of mechanical energy, and decreased rate of penetration, leading to an increase in non-productive time and drilling costs. Meanwhile, drillstring vibration monitoring using downhole sensors faces challenges, such as downhole data transmission and high sensor costs. Hence, this research employed machine learning techniques, namely, linear regression (LR), k-nearest neighbors (KNN), decision trees (DT), extreme gradient boosting (XGBoost), and random forest (RF), to monitor downhole vibrations using surface drilling parameters. The study passed through different phases: data collection, preprocessing, analysis, model development, model predictability evaluation, and model deployment. A dataset of 10,470 measurements was utilized to train and evaluate the developed models. The accuracy of these models in predicting downhole vibrations was assessed using statistical metrics such as root mean square error (RMSE) and coefficient of determination (R^2). The results of the training and testing phases showed a high degree of accuracy for the KNN, XGBoost, and RF models. Furthermore, the developed models were validated using an unseen dataset comprising 2,618 measurements, which confirmed the effectiveness of KNN, XGBoost, and RF models in the detection of downhole vibrations, with R^2 ranging from 0.88 to 0.96. Overall, this research shows the ability of the developed models for real-time prediction of downhole vibrations using surface drilling data without the need for expensive downhole sensors, resulting in reduced drilling costs.

Keywords: *Drillstring vibration; Surface drilling parameters; Machine learning; Real-time; Drilling optimization.*

1. Introduction

Downhole drillstring vibration is recognized as a significant issue encountered during drilling operations, leading to severe wear and tear to downhole equipment, frequent tool failures, a decrease in the rate of penetration (ROP), and an increase in the occurrence of non-productive time, which in turn results in escalated drilling costs. Additionally, it induces instability in the borehole and diminishes the amount of mechanical energy transferred to the drill bit [1-2]. Hence, the drilling team has a vital role during the drilling operation by diligently monitoring and diagnosing the drill string vibration signs to minimize any potential consequences [3]. The drilling process relies on the transmission of axial and rotational power to the drill bit, enabling it to penetrate the rock formation. Nevertheless, the bit's contact with the rocks, the interaction between the drillstring and the borehole walls, the eccentric rotation of the drillstring, and the occurrence of hole washouts may cause or exacerbate vibrations on the drillstring and the bit [4-5]. Downhole drillstring vibrations are commonly categorized into three main types: axial, lateral, and torsional vibrations. The axial vibration propagates along the drillstring due to upward and downward oscillations along its axis, resulting in the bit bounce phenomenon. This issue is particularly prevalent when drilling shallow vertical holes with large diameters, especially when utilizing tri-cone bits. Bit bounce can lead to detrimental effects such as damage to the downhole tools, drill bits, bottom hole assembly (BHA), and frequent fluctuations in the weight on bit (WOB). Consequently, ROP and overall drilling performance are negatively impacted [4,6]. Torsional vibration arises due to the non-uniform and restricted rotation of the drillstring, causing drillstring connection fatigue, leading to drillstring twist-off, damage to the bit and stabilizer cutters and gauges, and reduced drilling efficiency [7-8]. This type of vibration occurs when the torque required to rotate the drillstring and the bit is insufficient, causing the drillstring to momentarily stall (stick) until enough energy is regained to overcome resistance, allowing the drillstring to resume rotation (slip) at 2 to 15 times its average rotational speed [3,9]. This mechanism is known as stick-slip. Torsional vibration is prone to happen in highly deviated or extended-reach wells, hard formations, and deep wells. Another factor contributing to vibrations is the aggressive design of polycrystalline diamond compact (PDC) bits and the utilization of suboptimal drilling parameters like excessive WOB [4,10]. Lateral vibration is the third mode of drillstring vibration, in which the direction of vibration is perpendicular to the axis of the drillstring. The main source of lateral vibration is the drillstring rotation along an axis other than the wellbore axis, known as the whirl phenomenon [11]. These conditions can be initiated due to poor stabilization of BHA, hole washouts, and BHA resonance at critical rotational speeds [1,12]. The backward whirl is the most damaging vibration type, leading to destructive lateral shocks between the borehole and drillstring, which in turn results in downhole tool failures, rapid fatigue in drillstring and BHA connections, broken drill bit cutters, and enlarged boreholes [4,13]. Over the past years, several dynamic drillstring models and approaches have been established to optimize the drilling practice to be followed during downhole vibration occurrence and model drillstring vibration to diminish its associated consequences[14-15]. These approaches include modifying the drill bits' design and optimizing drilling parameters such as weight on bit, drillstring rotational speed (DRS), drilling torque (T), mud flow rate (Q), and rate of penetration [5,16]. The drilling team must follow a proficient mitigation practice for each mode of drillstring vibrations to increase drilling efficiency and minimize any negative consequences.

1.1. Machine learning and drilling engineering

Machine learning (ML) involves the study of computer algorithms that are designed to empower systems with the capability to learn autonomously through experience [17]. This field is widely acknowledged as a branch of artificial intelligence. Due to advancements in ML techniques, computers have acquired the ability to make decisions autonomously [18]. In recent years, machine learning and artificial intelligence have been extensively utilized in various applications within the oil and gas industry [19‑22]. The utilization of ML applications has a profound influence on drilling operations through the creation of exceptional models. These models have effectively minimized non-productive time and costs by addressing various aspects, including the rate of penetration prediction [23‑25], early detection of kicks [26-27], and the determination of rheological parameters of drilling fluids [28-29]. Additionally, ML applications have found relevance in numerous other areas within the drilling industry.

1.2. Machine learning techniques for downhole vibration detection

Recently, numerous sophisticated data analytics techniques and ML models have been developed to estimate and detect downhole vibrations. Baumgartner and van Oort [30] analyzed high-frequency downhole data via ML techniques to identify downhole vibrations. The model shows a success rate exceeding 90% when tested on high-frequency field datasets. Zhao *et al.* [16] utilized a variety of ML techniques in their study, including modified symbolic aggregate approximation, dynamic time wrapping, hierarchical clustering, pattern recognition, and classification, to analyze drilling data and identify abnormal drilling events such as stick-slip, whirling, stuck pipe, and loss of circulation. Zha and Pham [31] established an advanced deep learning model to analyze surface drilling parameters and predict the corresponding downhole stick-slip vibrations. The model's training and validation process involved the utilization of a dataset consisting of 1400 measurements of surface and downhole drilling data. Surface data included drilling torque, tension, rotary speed, WOB, and tri-axial acceleration, while downhole data consisted of torque, RPM, and acceleration. The model demonstrated a validation accuracy of 99% and a precision of 97%. Millan *et al.* [3] designed two models to detect and categorize the severity levels of stick-slip and lateral shocks using support vector machines (SVM) and bagged decision tree classification methods. The models were trained using surface drilling parameters such as WOB, T, hook load, DRS, and their associated vibration types and intensity levels. The stick-slip model was trained with 260 measurements, and the results showed an accuracy of 98% coefficient of determination. On the other hand, the lateral shocks model was trained with a larger dataset of 865 measurements and exhibited a slightly lower accuracy of 93% coefficient of determination. Ignova *et al.* [32] created a classification model using a fast Fourier transform (FFT), followed by two unsupervised ML techniques, a k-means cluster analysis algorithm, and a principal component analysis (PCA) pattern recognition technique. The model underwent a training process using 2048 measurements of high-frequency shock data to classify anomalies in drilling events effectively. Okoli *et al.* [2] employed five ML algorithms to classify downhole lateral and axial vibrations using primary surface drilling parameters T, ROP, and WOB. The results revealed that the classification ML for the intra-BHA runs achieved an accuracy level ranging from 50% to 80% in predicting the severity of downhole vibrations. However, the inter-BHA predictability experienced a notable reduction, which can be related to various operational events and wellbore conditions. Hegde *et al.* [10] developed a robust ML model to classify the severity of stick-slip in drilling operations. The stick-slip index, which measures the intensity of stick-slip caused by drilling vibrations, is categorized as either high or low through the utilization of ML classification algorithms like logistic regression, support vector machines, random forests, gaussian mixture models, and discriminant analysis. Training of the model involved input data from drilling operation parameters such as WOB, Q, RPM, ROP, and torque-on-bit. Among the various algorithms utilized, the random forest algorithm demonstrated superior performance with an accuracy of 90%. Gupta et al. [33] utilized the random forest and gradient boosting ML algorithms to develop a stick-slip classification model using drilling parameters such as T, ROP, and WOB as inputs. While the model demonstrates satisfactory performance for the low, medium, and high stick-slip classes, it exhibits subpar accuracy in predicting the severe class. The overall accuracy of the model is 62%, but the accuracy for the low stick-slip class is a high of 80%. Saadeldin *et al.* [34] developed a model to identify the three modes of vibration in real-time by employing an artificial neural network. The model evaluation demonstrated exceptional predictive capabilities, as indicated by statistical metrics that revealed a coefficient of correlation exceeding 0.95. Furthermore, the model exhibited minimal errors, with an average absolute percentage error of less than 3.5% when comparing actual and predicted values. Alsaihati and Alotaibi [35] employed K-nearest neighbors, logistic regression, and random forests machine learning techniques to assess the intensity of downhole lateral and torsional vibrations during surface hole drilling. A dataset comprising real-time field measurements from 10 wells was utilized to develop predictive models. The input variables consisted of surface drilling data and daily drilling reports, whereas the downhole vibrations encountered by these drilling parameters were used as output variables. The results revealed that the logistic regression model excelled in its precision in predicting the severity of downhole lateral and torsional vibrations, surpassing the performance of other models.

According to prior studies, the integrated detection of the three modes of vibration using ML research is currently insufficient. The main goal of this research is to establish comprehensive ML models for integrated prediction of drillstring vibration by utilizing surface drilling parameters, mud weight, and temperature obtained from rig sensors, allowing real-time detection of axial, lateral, and stick-slip vibrations. To accomplish this goal, five machine learning techniques, namely linear regression (LR), k-nearest neighbor (KNN), decision tree (DT), extreme gradient boosting (XGBoost), and random forest (RF), have been employed to generate and evaluate the accuracy of these models for the prediction of downhole vibrations. The outcomes of this research will empower the drilling crew to frequently identify drillstring vibrations and apply the appropriate drilling practices to mitigate their consequences, reducing non-productive time and supporting drilling operations optimization.

2. Methodology

2.1. Study approach

The process of developing machine learning models for predicting downhole vibrations involves a systematic process that starts with gathering the data, data preprocessing and analysis, constructing the model, and model performance evaluation before finally deploying the model and saving the most optimal outcomes for vibration predictions. Figure 1 depicts the sequential flowchart of the steps involved in developing the models.

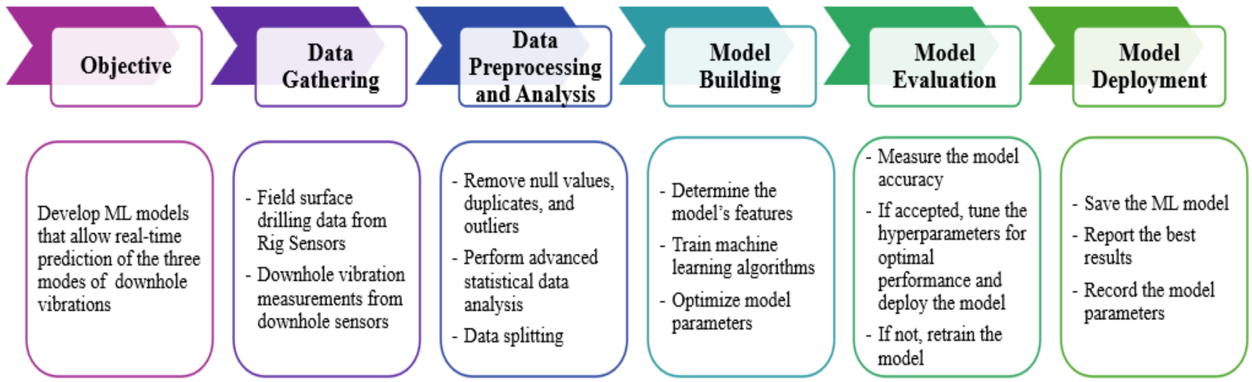

Figure 1. ML model development workflow diagram.

2.2. Data collection and preparation

Field drilling data was gathered from five wells located in the western desert of Egypt. The data was collected while drilling a 12 ¼" intermediate hole to eliminate the effect of hole size change. These five wells were selected from two neighboring fields to minimize the influence of formation lithology and mineral composition variations. The 12 ¼" hole in the selected wells was drilled using 6 blades and 16 mm cutter-size PDC bits with identical bit designs, resulting in a negligible impact on the bit-lithology interaction. Additionally, the intermediate section of these wells was drilled using water-based mud with consistent mud properties.

The collected drilling data consists of surface drilling parameters employed as inputs in the model, along with their corresponding downhole vibration measurements, which serve as the desired output of the ML model. Several rig sensors are employed to measure these surface drilling parameters, including weight on bit (WOB), drillstring rotational speed (DRS) in revolutions per minute (RPM), drilling torque (T), mud flow rate (Q), standpipe pressure (SPP), rate of penetration (ROP), as well as hole depth and other drilling fluid properties like mud weight (M.Wt) and mud temperature (M.T). Downhole BHA sensors are used to measure downhole drillstring vibration parameters such as lateral vibration, axial vibration, stick-slip percentage (SS%), and vibration count (CPS). The vibration count indicates the number of vibration shocks that occurred per second. The stick-slip percentage represents the severity of torsional vibrations and is computed from the drill bit rotation sensor [36]. The stick-slip percentage is calculated as follows:

$$SS\% = \frac{|\Delta\ Bit\ RPM|}{Average\ Bit\ RPM} \qquad (1)$$

Adequate data preprocessing plays a crucial role in improving the efficiency of machine learning models. Since the quality of the data used is a critical factor for ML models to detect physical phenomena and the presence of noise within a dataset has the potential to impact the accuracy of predictions made by ML algorithms [37-38], the collected data was preprocessed using techniques like Z-score and box plots. This preprocessing phase eliminated outliers, missing data, unreasonable sensor readings, duplicates, and negative values, thereby enhancing the overall performance of ML algorithms. Figure 1 and Figure 3 show lateral vibration values before and after outliers were removed using the box plot technique.

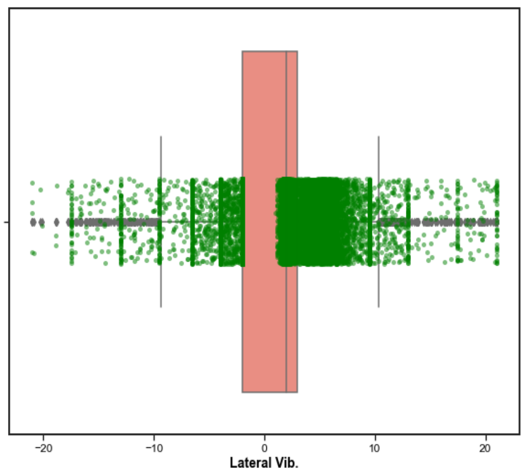

Figure 1. Lateral vibration values before removing outliers.

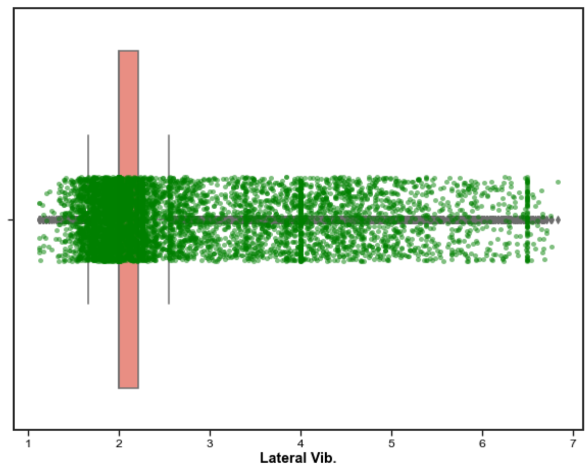

Figure 3. Lateral vibration values after removing outliers.

After the collected data was directed to the cleaning and preprocessing phase, a dataset comprising 10470 data points was obtained. The dataset contained ten input variables and four output variables. Each ML model underwent training with 70% of the data points, while the remaining 30% were allocated for model evaluation. Subsequently, the trained models underwent a further testing phase using an unseen dataset of 2618 data points obtained from a nearby well while drilling a 12 ¼'' intermediate hole to validate the model prediction capability for new wells.

2.3. Data analysis

The preprocessed data was subjected to statistical analysis, as shown in Table 1. Furthermore, Figure 4 shows the density plots that were utilized to visually represent the numeric distribution of each parameter within the dataset.

The extensive range of values for the parameters, as indicated by the maximum and minimum values, is beneficial in improving the efficacy of the model training phase. The hole depth ranged from 1146 to 3002.8 m, the weigh on bit from 1.5 to 55.8 Klb, the drillstring rotational speed from 56.5 to 147.6 RPM, the drilling torque from 0.92 to 16.98 Klb.ft, the mud flow rate from 580.8 to 860 gallons per minute (GPM), the standpipe pressure from 1560.4 to 3187.6 psi, the rate of penetration from 2.7 to 88 m/hr, the mud weight from 9 to 9.6 ppg, and the mud temperature from 37.2 to 74.2 °C. The associated downhole vibrations ranged from 1.1 to 6.8 g, 0.5 to 2.9 g, 1.5 to 249.1 %, and 2 to 208.6 CPS for lateral vibration, axial vibration, stick-slip percentage, and vibration count respectively. Figure 5 illustrates the pair plots for the dataset features to visualize the relationship between each other. The pair plot showed that the correlation between the model parameters is nonlinear and very complex. A correlation matrix was employed to determine the correlation coefficient among all input and output model variables, as illustrated in Figure 2. The correlation matrix demonstrated the complicated relationships among model parameters, emphasizing the importance of employing machine learning techniques to model these parameters and identify the interrelationships between them accurately.

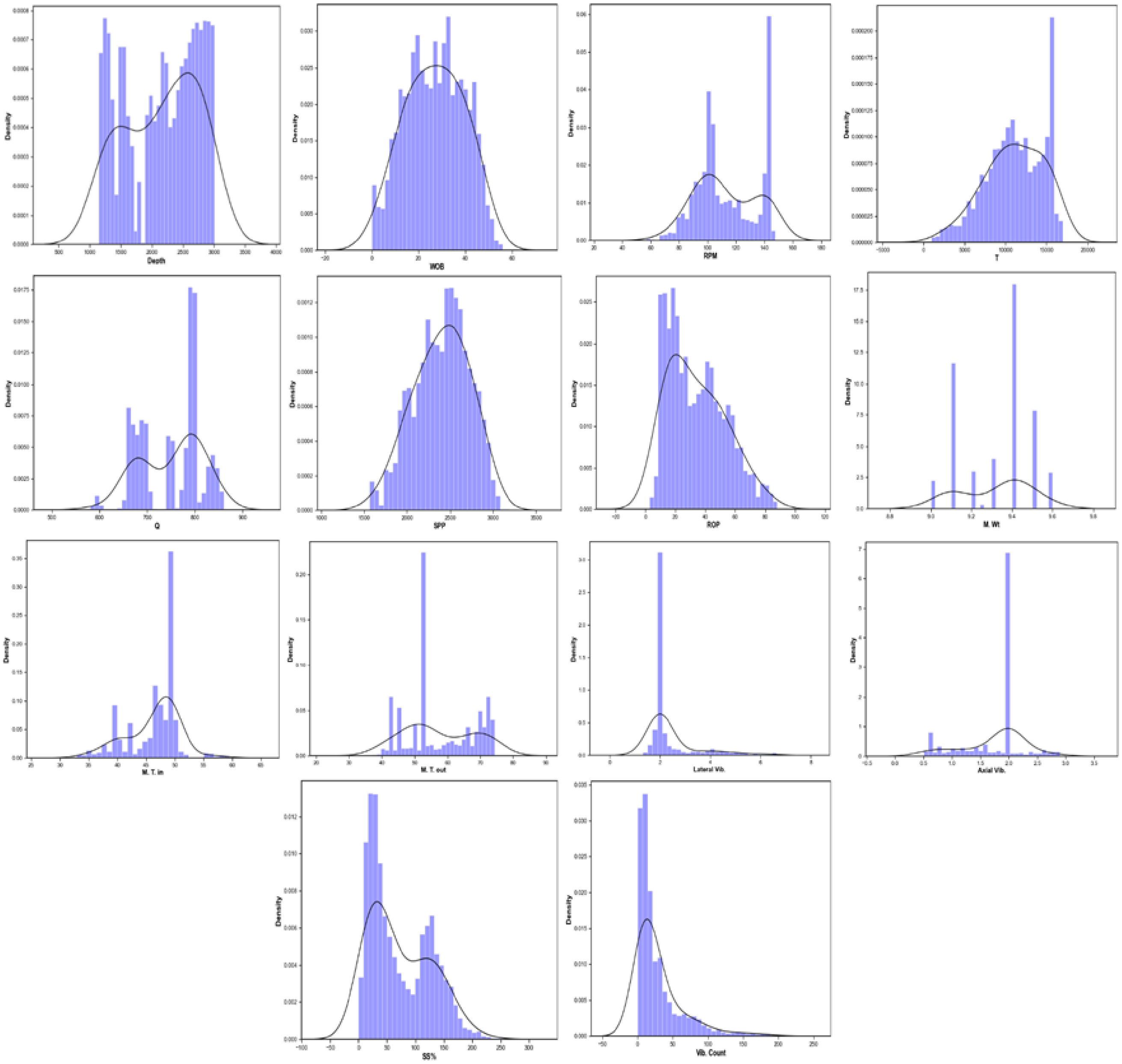

Figure 4. Density plot for each parameter in the dataset after the preprocessing phase.

Table 1 Data statistical analysis.

SP	Depth (m)	WOB (Klb)	DRS (RPM)	T (lb.ft)	Q (GPM)	SPP (psi)	ROP (m/hr)	M.Wt (ppg)	M.T. in (°C)	M.T. out (°C)	LV. (g)	AVib. (g)	SS% (%)	VC (CPS)
Min	1146.0	1.5	56.5	922.1	580.8	1560.4	2.7	9.0	32.8	37.2	1.1	0.5	1.5	2.0
25%	1566.2	17.4	98.3	8641.4	686.3	2176.9	17.8	9.1	43.6	51.5	2.0	1.5	26.6	7.6
50%	2228.6	27.1	106.4	11146.7	777.8	2431.3	30.9	9.4	47.7	52.1	2.0	2.0	54.8	15.4
75%	2669.8	36.8	137.7	14178.6	794.9	2635.9	47.8	9.4	49.2	68.6	2.2	2.0	120.3	35.7
Max	3002.8	55.8	147.6	16979.2	860.0	3187.6	88.0	9.6	59.4	74.2	6.8	2.9	249.1	208.6
Av	2152.1	26.8	112.9	11058.6	749.9	2399.4	34.1	9.3	46.3	57.3	2.4	1.8	72.9	29.0
Std	580.7	12.4	20.9	3511.6	63.4	317.8	19.0	0.2	4.4	10.4	1.0	0.5	53.1	33.6

SP- Statistical parameter; Av- average; LV- lateral vibration; AVib-axial vibration; VC- vibration count

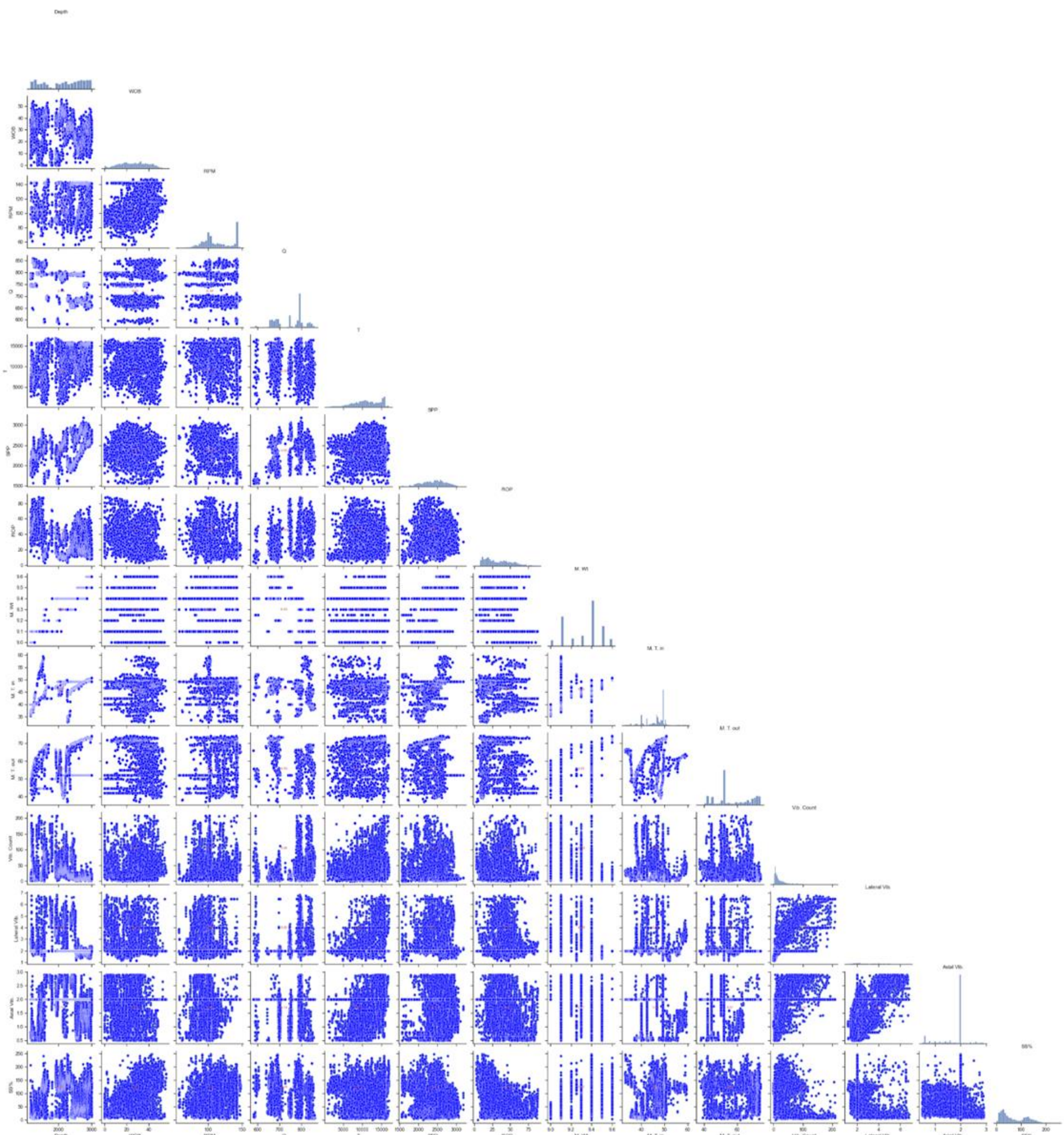

Figure 5. Pair plots for dataset features.

2.4. Model development

Machine learning techniques have an extensive range of applications in the oil and gas sector, especially in drilling operations. One of these applications is the real-time detection of downhole drillstring vibrations, this aids drilling engineers in adapting the drilling parameters to eliminate the consequences caused by these vibrations. Machine learning algorithms such as Linear Regression, k-nearest neighbor, decision tree, extreme gradient boosting, and random forest were employed to train models and evaluate their capability for downhole vibration prediction.

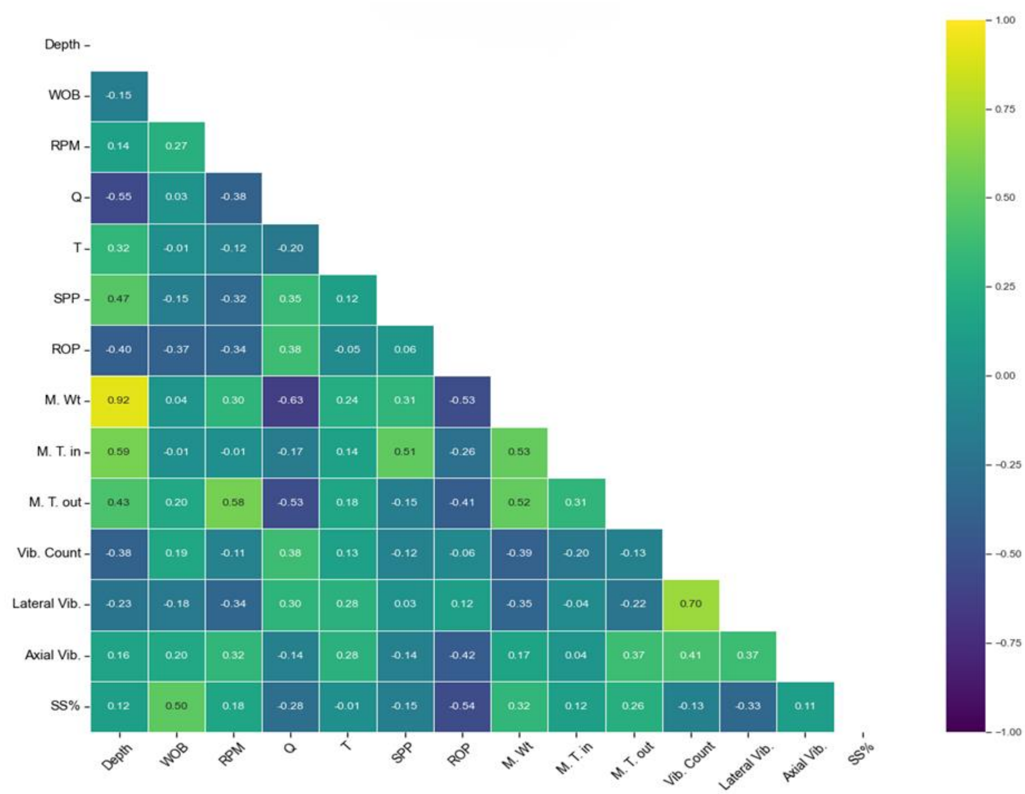

Figure 2. Correlation matrix for model parameters.

Linear regression is a popular machine learning approach that predicts continuous numerical values. It seeks to create a linear correlation between a dependent variable and one or more independent variables [39]. The model assumes that the relationship between the variables may be represented by a straight line, making it a simple but effective tool for prediction and analysis. One of LR's most important strengths is its interpretability. The model provides a linear equation that directly links the vibration response (dependent variable) to a collection of drilling parameters (independent variables) that are thought to impact vibrations. However, LR's fundamental drawback is its assumption of linearity. Drill string vibrations are frequently affected by complicated, non-linear interactions among drilling factors. As a result, LR models may fail to capture the entire range of vibration behavior. This constraint requires the development of alternative machine learning approaches capable of handling such complexities.

K-nearest neighbor is an instance-based or non-generalizing algorithm that is applicable for both regression and classification tasks. It is a non-parametric algorithm that relies on a similarity measure, such as distance functions like Euclidean or Hamming distance. Initially introduced by Fix and Hodges in 1951, this method is now commonly known as the k-nearest neighbors method [40]. The K-NN algorithm is widely regarded as one of the simplest ML algorithms, and it is very effective when dealing with large training datasets [35]. KNN stores all training samples in memory, unlike alternative learning algorithms that allow discarding the training data after the model is built [41]. Many parameters have a direct effect on the performance of the KNN model; however, the two most crucial ones are the number of nearest neighbors (n_neighnors) and the weights function. In this study, the number of nearest neighbors is set equal to 5, and the weights function is used as ''uniform'', indicating that every point within the neighborhood is given equal weight.

Decision tree models provide a basic and understandable method for machine learning. They partition the feature space using simple decision rules, making them useful for determining the link between drilling parameters and vibration dynamics in the drill string [2]. Despite their simplicity, decision trees have the capability to effectively handle a diverse array of data and provide insight into the importance of features. However, they are prone to overfitting, especially when dealing with noisy or multidimensional data [42]. Many parameters affect the performance of the decision tree model, such as the maximum depth of the tree (max_depth), which is used as ''none'', the strategy used to choose the split at each node (splitter), which is used as ''best'', and the minimum number of samples required to split an internal node (min_samples_split), which is set equal to 2.

Extreme gradient boosting is a machine learning technique like having a team of experts work together. Each "expert" learns from the mistakes of the others, resulting in a super-accurate prediction of vibrations. XGBoost, an extension of gradient boosting, has received great praise for its superior predictive capability and adaptability. Its distinguishing features, including innovative regularization algorithms and efficient parallel processing, make it capable of handling complicated, high-dimensional datasets [43]. In terms of drill string vibration prediction, XGBoost's capacity to describe complex interactions between drilling parameters and vibration dynamics makes it an appealing option. Its interpretability increases its utility by providing useful insights into the underlying causes of vibrations.

Random forest is an ensemble-based machine learning algorithm that performs well for low-dimensional drilling data [20]. It is a type of meta-estimator. It works by fitting multiple decision tree regressors on different parts of the dataset and then averaging their predictions. This technique helps enhance predictive accuracy and prevent overfitting. Trees in the forest use the best-split strategy. Random forests are highly effective when dealing with noise, multi-attribute data, and tuning algorithms [44]. The RF model has different parameters, such as the number of trees in the forest (n-estimators), which is set equal to 100; the minimum number of samples required to split an internal node (min_samples_split), which is set equal to 2; and the minimum number of samples needed to be at a leaf node (min_samples_split), which is set equal to 1.

As the interpretation of ML models is very complex, it is essential to utilize explanatory ML methods to determine the primary features influencing the vibration model's output. The SHAP (SHapley Additive exPlanations) technique is one of these methods derived from game theory, plays a crucial role in enhancing the clarity and interpretability of ML models, and discloses the significance of each input feature on the predicted vibration values [45]. The SHAP value provides insight into the extent to which a feature value has influenced the prediction of a specific instance in relation to the average prediction of the dataset. The shapely values are demonstrated through the utilization of waterfall and local bar plots, as shown in Figure 7, where the length of each bar corresponds to the SHAP value of a particular feature. The plots indicate that depth, mud temperature, WOB, and T are the main features that affect the vibration model prediction.

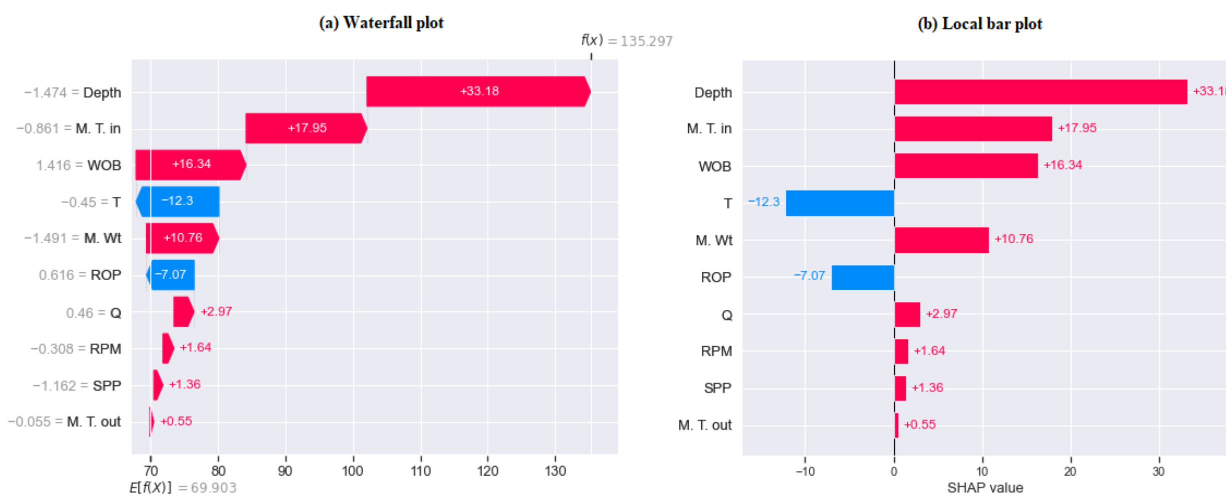

Figure 7. Feature importance using SHAP values, represented by the waterfall plot and the local bar plot.

2.5. Model evaluation

The accuracy of each model was evaluated through a statistical analysis of their respective errors. This analysis involved measuring the root mean square error (RMSE) and coefficient of determination (R^2) values between the actual field measurements for vibrations and the predicted vibration values generated by the models. These statistical parameters, as defined in equations (2) and (3), were utilized to quantify the predictive capability of the models. A smaller RMSE value indicates a higher level of accuracy in downhole vibration prediction by the developed models. Furthermore, the R^2 value ranges from 0 to 1, with a higher value approaching 1, indicating a stronger match between the model's predictions and the actual data points.

Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_p\right)^2}{n}} \quad (2)$$

Coefficient of determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_p\right)^2}{\sum_{i=1}^{n}\left(y_i - \left(\frac{1}{n}\sum_{i=1}^{n} y_p\right)\right)^2} \quad (3)$$

where n is the number of data points; y_i is the actual value of data point I; and y_p is the predicted value returned from the model.

3. Results and discussion

The developed ML models were trained using 7329 data points, which accounted for 70% of the dataset. Subsequently, 3141 data points, representing 30% of the dataset, were employed for testing these models to evaluate the models' performance in real-time downhole vibration prediction.

The training phase's evaluation parameters for developed models are presented and measured in Figure 8 and Table 2. It is obvious that the DT model overfits the training data, indicating that the model may have excessively learned the training data. The LR model exhibited poor performance, as indicated by an RMSE of 22.89 and an R^2 value of 0.42. Conversely, the KNN, XGBoost, and RF models demonstrated outstanding performance in predicting downhole vibration during the training phase, achieving R^2 values of 0.94, 0.98, and 0.99, respectively.

The testing phase of the models demonstrated great levels of accuracy for the XGBoost and RF models, as indicated in Figure 9 and Table 2. The performance of the LR model was not satisfactory, with an RMSE value of 23.17 and an R^2 value of 0.40. The KNN model showed higher accuracy than the LR, with an RMSE value of 12.20 and an R^2 value of 0.86. XGBoost and RF Models showed exceptional accuracy in integrated prediction of the downhole vibration, with an R^2 higher than 0.9. The RF model outperformed the other models with an RMSE value of 9.05 and an R^2 value of 0.93.

The developed models were subjected to an additional testing step (validation phase) to evaluate the accuracy of their predictability using a blind dataset of 2618 data points collected from a nearby well. Figure 10shows the high degree of match between the actual vibration measurement of the unseen dataset and the predicted values, with R^2 ranging from 0.88 to 0.96, as detailed in Table 2. Among the models, the RF model demonstrated the highest level of accuracy in real-time prediction of downhole vibration measurements.

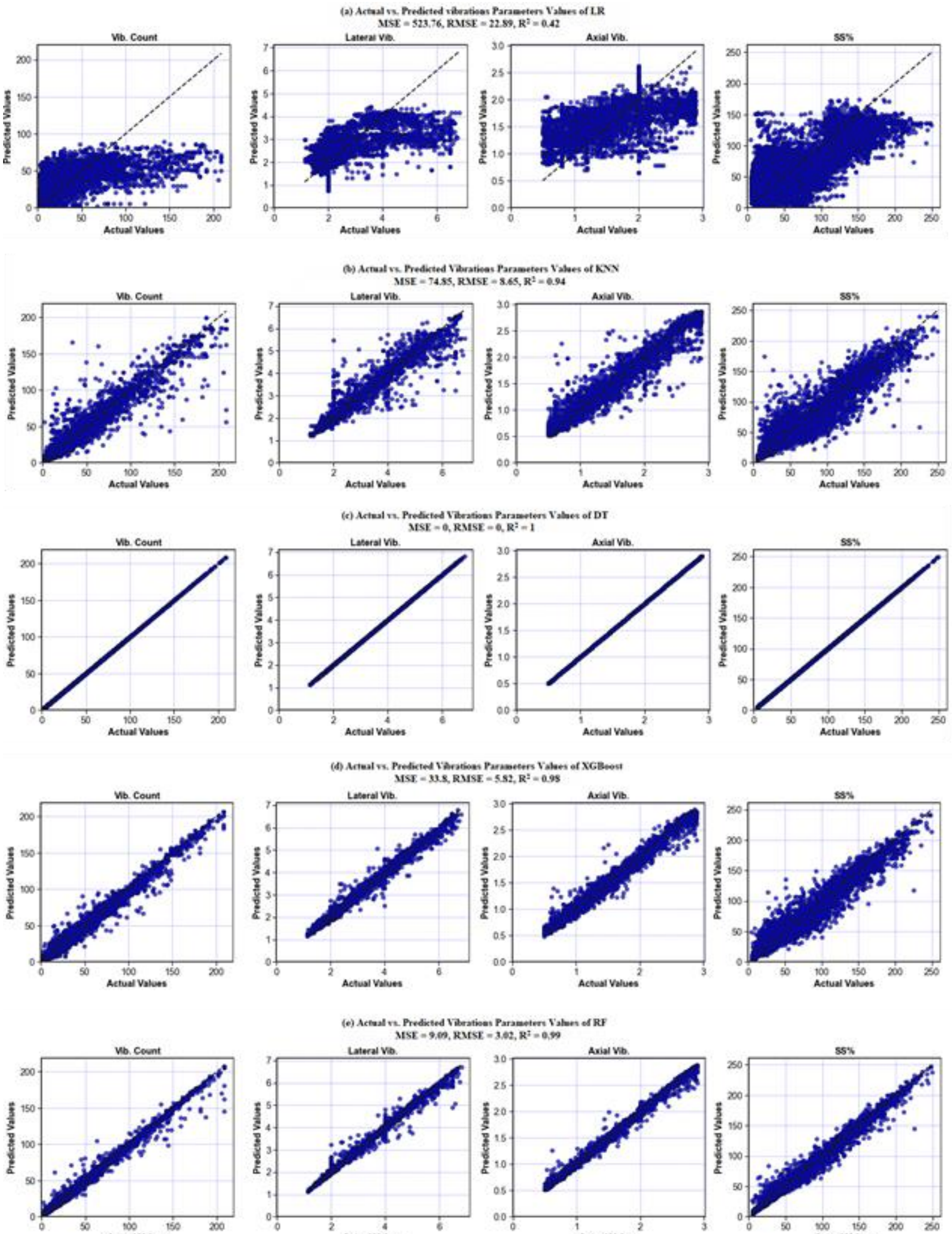

Figure 8. Actual vibration test data versus the predicted one during the training phase of the developed vibration models, where (a) is for LR, (b) is for KNN, (c) is for DT, (d) is for XGBoost, and (e) is for RF models.

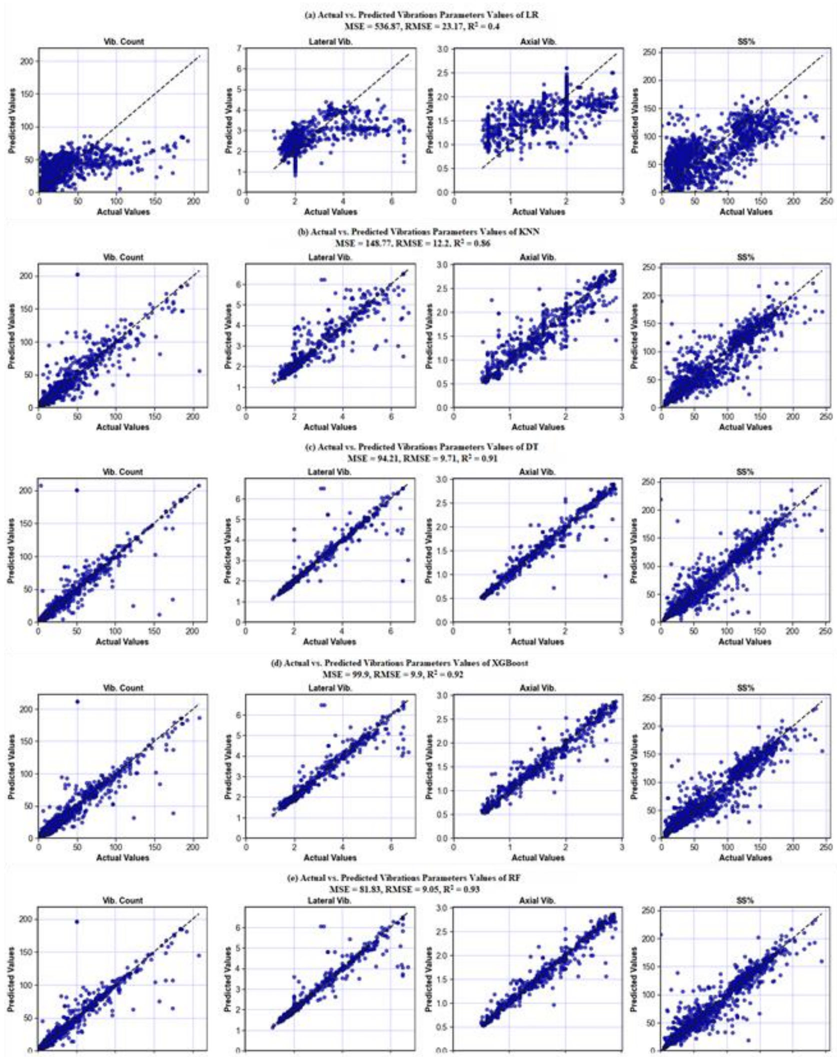

Figure 9. Actual vibration test data versus the predicted one during the testing phase of the developed vibration models, where (a) is for LR, (b) is for KNN, (c) is for DT, (d) is for XGBoost, and (e) is for RF models.

Table 2. Results of developed models.

Developed models	Dataset	Accuracy measurement		
		MSE	RMSE	R^2
LR	Training	523.76	22.89	0.42
	Testing	536.87	23.17	0.4
	Validation	527.94	22.98	0.42
KNN	Training	74.85	8.65	0.94
	Testing	148.77	12.2	0.86
	Validation	129.93	11.31	0.88
DT	Training	0	0	1
	Testing	94.21	9.71	0.91
	Validation	69.13	8.31	0.95
XGBoost	Training	33.8	5.82	0.98
	Testing	99.9	9.9	0.92
	Validation	79.26	8.9	0.95
RF	Training	9.09	3.02	0.99
	Testing	81.83	9.05	0.93
	Validation	60.89	7.8	0.96

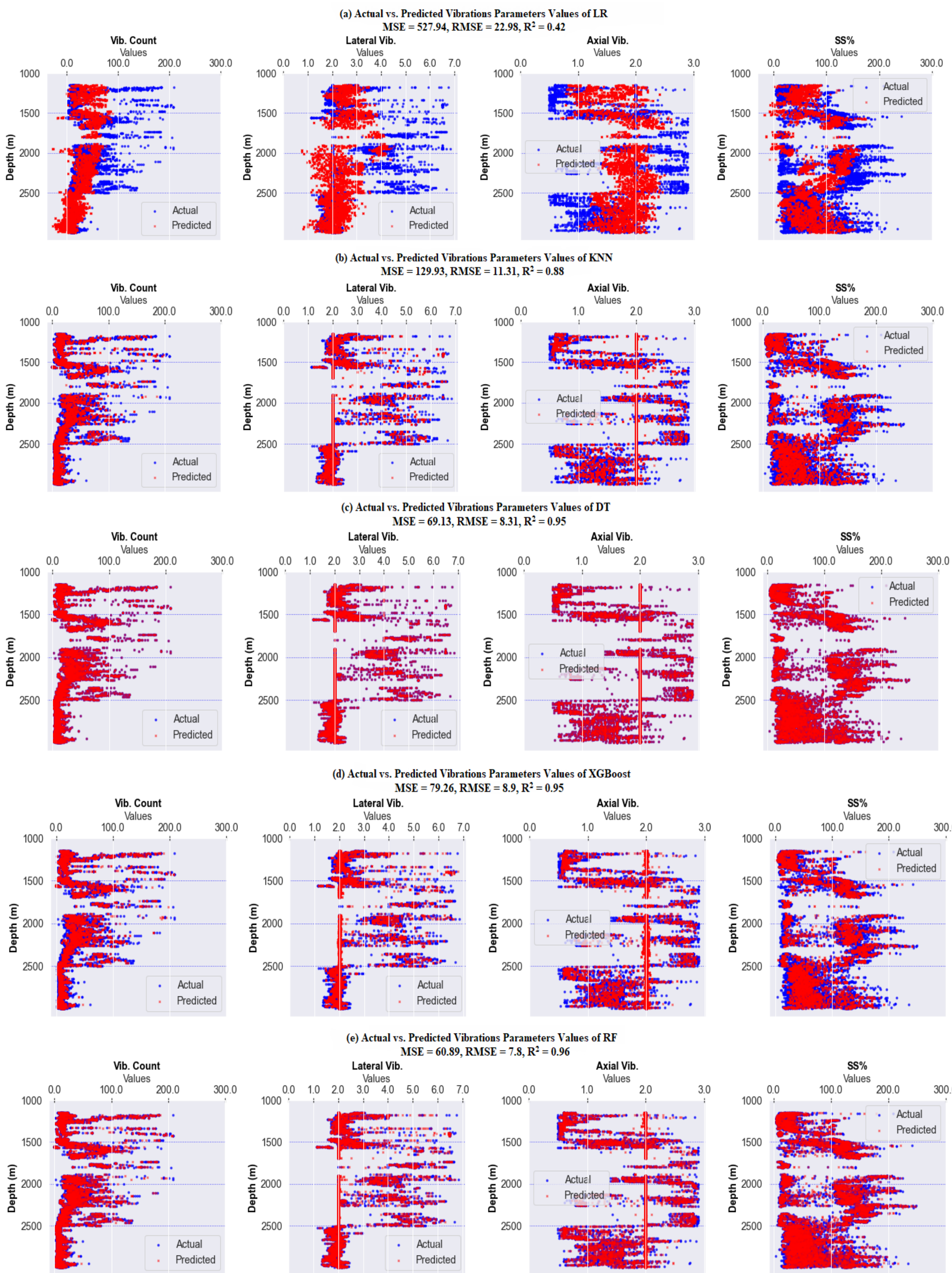

Figure 10. Actual vibration test data versus the predicted one during the validation phase of the developed vibration models, where (a) is for LR, (b) is for KNN, (c) is for DT, (d) is for XGBoost, and (e) is for RF models.

4. Conclusion

This study has developed various machine learning models to predict different types of downhole vibrations, like lateral, axial, and stick-slip, and to identify their specific characteristics. The models were developed using only surface drilling data as their input. The machine learning models were trained and tested using data gathered while drilling a 12 ¼″ hole. This data included surface drilling parameters along with their corresponding downhole vibrations.

Downhole vibration prediction significantly impacts the optimization of drilling operations as it allows for the implementation of appropriate drilling parameters, prevents downhole tool failures, and minimizes non-productive time. A dataset comprising 10470 data points was utilized to train and test LR, KNN, DT, XGBoost, and RF ML algorithms for integrated prediction of downhole vibrations using only surface drilling data. Overfitting was observed in the decision tree model during the training phase. The results showed that XGBoost and RF models have a high degree of accuracy in predicting downhole vibration parameters, with an R^2 value greater than 0.92 and an RMSE less than 10 during the training and testing phases.

The developed models underwent another testing step with an unseen dataset to check the performance of these models in downhole vibration prediction, and the accuracy recorded showed R^2 higher than 0.95 for the XGBoost and RF models. The LR model exhibited the poorest performance compared to the other developed models during the training and evaluation phases, with an R^2 value of 0.42 during the validation phase. The RF model showed superior accuracy in downhole vibration prediction, surpassing the other models with an R^2 value of 0.96 and an RMSE of 7.8.

Nomenclature

BHA	*Bottom Hole Assembly*	*Q*	*Flow Rate*
CPS	*Count Per Second*	*R^2*	*Coefficient of Determination*
DRS	*Drillstring Rotational Speed*	*RF*	*Random Forest*
DT	*Decision Tree*	*RMSE*	*Root Mean Square Error*
GPM	*Gallons per Minute*	*ROP*	*Rate of Penetration*
KNN	*k-nearest neighbor*	*RPM*	*Revolutions Per Minute*
LR	*Linear Regression*	*SHAP*	*SHapley Additive exPlanations*
M.T	*Mud Temperature*	*SPP*	*Standpipe Pressure*
M.Wt	*Mud Weight*	*SS%*	*Stick-Slip Percentage*
ML	*Machine learning*	*T*	*Drilling Torque*
MSE	*Mean Square Error*	*WOB*	*Weight on Bit*
PDC	*Polycrystalline Diamond Compact*	*XGBoost*	*Extreme Gradient Boosting*

Acknowledgment

Sincere appreciation is extended to all staff members of the Petroleum Engineering Department for their valuable support and motivation throughout this research work, as well as for providing an excellent research environment.

References

[1] Allen MB. BHA Lateral Vibrations: Case Studies and Evaluation of Important Parameters. Paper presented at the SPE/IADC Drilling Conference, New Orleans, Louisiana, March 1987 (pp. SPE-16110). https://doi.org/10.2118/16110-MS

[2] Okoli P, Vega JC, and Shor R. Estimating Downhole Vibration via Machine Learning Techniques Using Only Surface Drilling Parameters. Paper presented at the SPE Western Regional Meeting, San Jose, California, USA, April 2019 (pp. SPE-195334). https://doi.org/10.2118/195334-MS

[3] Millan E, Ringer M, Boualleg R, and Li D. Real-Time Drillstring Vibration Characterization Using Machine Learning. Paper presented at the SPE/IADC International Drilling Conference and Exhibition, The Hague, The Netherlands, March 2019 (pp. SPE-194061). https://doi.org/10.2118/194061-MS

[4] Sotomayor, G. P. G.; Placido, J. C.; Cunha, J. C. Drill String Vibration: How to Identify and Suppress. In All Days; SPE, 1997. https://doi.org/10.2118/39002-MS
[5] Hegde C, Pyrcz M, Millwater H, Daigle H, and Gray K. Fully Coupled End-to-End Drilling Optimization Model Using Machine Learning. J Pet Sci Eng, 2020; 186, 106681. https://doi.org/10.1016/J.PETROL.2019.106681
[6] Zhu Q, Zou Z, Huang B, Ma L, and Xia J. Downhole Vibration Causing a Drill Collar Failure and Solutions. Natural Gas Industry B, 2017; 4(2): 73-80. https://doi.org/10.1016/j.ngib.2017.07.013
[7] Osnes SM, Amundsen PA, Weltzin T, Nyrnes E, Hundstad BL, and Grindhaug G. MWD Vibration Measurements: A Time for Standardisation. Paper presented at the SPE/IADC Drilling Conference and Exhibition, Amsterdam, The Netherlands, March 2009 (pp. SPE-119877). https://doi.org/10.2118/119877-MS
[8] Kyllingstad A, and Halsey GW. A Study of Slip-Stick Motion of the Bit. Paper presented at the SPE Annual Technical Conference and Exhibition, Dallas, Texas, September 1987 (pp. SPE-16659). https://doi.org/10.2118/16659-MS
[9] Shen Y, Zhang Z, Zhao J, Chen W, Hamzah M, Harmer R, and Downton G. The Origin and Mechanism of Severe Stick-Slip. Paper presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, October 2017 (pp. SPE-187457). https://doi.org/10.2118/187457-MS
[10] Hegde C, Millwater H, and Gray K. Classification of Drilling Stick Slip Severity Using Machine Learning. J Pet Sci Eng, 2019; 179: 1023-1036. https://doi.org/10.1016/j.petrol.2019.05.021
[11] Brett JF, Warren TM, and Behr SM. Bit Whirl: A New Theory of PDC Bit Failure. Paper presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, October 1989 (pp. SPE-19571). https://doi.org/10.2118/19571-MS
[12] Minett-Smith D, Stroud D, and Pagett J. Real-Time Whirl Detector Aids Drilling Optimization. Paper presented at the SPE Annual Technical Conference and Exhibition, Florence, Italy, September 2010 (pp. SPE-135110). https://doi.org/10.2118/135110-MS
[13] Wang C, Li X, Li Y, Xu W, and Liao W. Analysis of the Effect of Whirl on Drillstring Fatigue. Shock and Vibration, 2021. https://doi.org/10.1155/2021/6666767
[14] Jogi PN, Macpherson JD, and Neubert M. Field Verification of Model-Derived Natural Frequencies of a Drill String. J Energy Resour Technol, 2002; 124(3): 154–162. https://doi.org/10.1115/1.1486018
[15] Aslaksen H, Annand M, Duncan R, Fjaere A, Paez L, and Tran U. Integrated FEA Modeling Offers System Approach to Drillstring Optimization. Paper presented at the IADC/SPE Drilling Conference, Miami, Florida, USA, February 2006 (pp. SPE-99018). https://doi.org/10.2118/99018-MS
[16] Zhao J, Shen Y, Chen W, Zhang Z, and Johnston S. Machine Learning-Based Trigger Detection of Drilling Events Based on Drilling Data. Paper presented at the SPE Eastern Regional Meeting, Lexington, Kentucky, USA, October 2017 (pp. SPE-187512). https://doi.org/10.2118/187512-MS
[17] Sayani JKS, and Lal B. Machine Learning in Oil and Gas Industry. In Machine Learning and Flow Assurance in Oil and Gas Production. Springer Nature Switzerland: Cham, 2023; pp 7–26. https://doi.org/10.1007/978-3-031-24231-1_2
[18] Sah S. Machine Learning: A Review of Learning Types. 2020. https://doi.org/10.20944/preprints202007.0230.v1
[19] Sircar A, Yadav K, Rayavarapu K, Bist N, and Oza H. Application of Machine Learning and Artificial Intelligence in Oil and Gas Industry. Petroleum Research, KeAi Publishing Communications Ltd, 2021; 6(4): 379-391. https://doi.org/10.1016/j.ptlrs.2021.05.009
[20] Hegde C, and Gray KE. Use of Machine Learning and Data Analytics to Increase Drilling Efficiency for Nearby Wells. J Nat Gas Sci Eng, 2017; 40: 327-335. https://doi.org/10.1016/J.JNGSE.2017.02.019
[21] Thabet S, Zidan H, Elhadidy A, Taman A, Helmy A, Elnaggar H, and Yehia T. Machine Learning Models to Predict Production Rate of Sucker Rod Pump Wells. Paper presented at the SPE Western Regional Meeting, Palo Alto, California, USA, April 2024 (pp. SPE-218857). https://doi.org/10.2118/218857-MS
[22] Barghoth TM, Elramah SG, Salem AM. Article Open Access Pore and Fracture Pressures Real Time Prediction during Drilling Operation Using a Regression Ensemble Machine Learning Model. Petroleum and Coal, 2024; 66(4): 1266–1282.

[23] El-Sayed Y, Salem AM, and El-Rammah S. Rate of Penetration Prediction in Drilling Operation in Oil and Gas Wells by K-Nearest Neighbors and Multi-Layer Perceptron Algorithms. Journal of Mining and Environment, 2023; 14(3): 755-770. https://doi.org/10.22044/jme.2023.12694.2306

[24] Yossef MA, and Farahat MS. A Novel Rate of Penetration Prediction Model for Hybrid Bit in Highly Unconfined Compressive Strength Formation. Petroleum and Coal, 2023; 65(3): 924-932.

[25] Shokry A, Elkatatny S, and Abdulraheem A. Real-Time Rate of Penetration Prediction for Motorized Bottom Hole Assembly Using Machine Learning Methods. Sci Rep, 2023; 13(1). https://doi.org/10.1038/s41598-023-41782-2

[26] Muojeke S, Venkatesan R, and Khan F. Supervised Data-Driven Approach to Early Kick Detection during Drilling Operation. J Pet Sci Eng, 2020; 192, 107324. https://doi.org/10.1016/j.petrol.2020.107324

[27] Chi P, Qingfeng L, Jianhong F, Yun Y, Xiaomin Z, Yu S, Zhaoyang X, Chengxu Z, and Pengcheng W. An Intelligent Model for Early Kick Detection Based on Cost-Sensitive Learning. Process Safety and Environmental Protection, 2023; 169: 398-417. https://doi.org/10.1016/j.psep.2022.10.086

[28] Davoodi S, Mehrad M, Wood DA, Ghorbani H, and Rukavishnikov VS. Hybridized Machine-Learning for Prompt Prediction of Rheology and Filtration Properties of Water-Based Drilling Fluids. Eng Appl Artif Intell, 2023; 123, 106459. https://doi.org/10.1016/j.engappai.2023.106459

[29] Alsabaa A, Gamal H, Elkatatny S, and Al Shehri DA. Rheology Predictive Model Based on an Artificial Neural Network for Micromax Oil-Based Mud. Arab J Sci Eng, 2023; 48(7): 9179-9193. https://doi.org/10.1007/s13369-022-07123-7

[30] Baumgartner T, Van Oort E. Pure and Coupled Drill String Vibration Pattern Recognition in High Frequency Downhole Data. Paper presented at the SPE Annual Technical Conference and Exhibition, Amsterdam, The Netherlands, October 2014 (pp. SPE-170955). https://doi.org/10.2118/170955-MS

[31] Zha Y, and Pham S. Monitoring Downhole Drilling Vibrations Using Surface Data through Deep Learning. In SEG Technical Program Expanded Abstracts; Society of Exploration Geophysicists, 2018; pp 2101-2105. https://doi.org/10.1190/segam2018-2964198.1

[32] Ignova M. Recognizing Abnormal Shock Signatures During Drilling with Help of Machine Learning. Paper presented at the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, March 2019 (pp. SPE-194952). https://doi.org/10.2118/194952-MS

[33] Gupta S, Chatar C, and Celaya JR. Machine Learning Lessons Learnt in Stick-Slip Prediction. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2019 (pp. SPE-197584). https://doi.org/10.2118/197584-MS

[34] Saadeldin R, Gamal H, Elkatatny S, and Abdulraheem A. Intelligent Model for Predicting Downhole Vibrations Using Surface Drilling Data during Horizontal Drilling. Journal of Energy Resources Technology, Transactions of the ASME, 2022; 144(8). https://doi.org/10.1115/1.4052794

[35] Saihati A. Determining Severity of Lateral and Torsional Downhole Vibrations While Drilling Surface Holes Using Three Machine Learning Techniques. Article in SPE Journal, 2022; 27(3): 1493-1503. https://doi.org/10.2118/209575-PA/2612919/spe-209575-pa.pdf

[36] Sugiura J, and Jones S. Real-Time Stick-Slip and Vibration Detection for 8 ½"-Hole-Size Rotary Steerable Tools in Deeper Wells and More Aggressive Drilling. Paper presented at AADE National Technical Conference and Exhibition, Houston, Texas, 2007 (pp. AADE-07-NTCE-47).

[37] García S, Luengo J, and Herrera F. Data Preprocessing in Data Mining. Springer International Publishing: Cham, 2015; 72. https://doi.org/10.1007/978-3-319-10247-4

[38] Gupta S, and Gupta A. Dealing with Noise Problem in Machine Learning Data-Sets: A Systematic Review. Procedia Comput Sci, 2019; 161: 466-474. https://doi.org/10.1016/j.procs.2019.11.146

[39] Lim HI. A Linear Regression Approach to Modeling Software Characteristics for Classifying Similar Software. In Proceedings - International Computer Software and Applications Conference, IEEE Computer Society, 2019; 1: 942-943. https://doi.org/10.1109/COMPSAC.2019.00152

[40] Swamynathan M. Mastering Machine Learning with Python in Six Steps. Apress: Berkeley, CA, 2019. https://doi.org/10.1007/978-1-4842-4947-5

[41] Vemuri VK. The Hundred-Page Machine Learning Book. Journal of Information Technology Case and Application Research, 2020; 22(2): 136-138. https://doi.org/10.1080/15228053.2020.1766224

[42] Kotsiantis SB. Decision Trees: A Recent Overview. Artif Intell Rev, 2013; 39(4): 261-283. https://doi.org/10.1007/s10462-011-9272-4

[43] Chen T, Guestrin C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM: New York, USA, 2016; pp 785-794. https://doi.org/10.1145/2939672.2939785

[44] Hastie T, Tibshirani R, and Friedman J. The Elements of Statistical Learning. Springer New York, 2009. https://doi.org/10.1007/978-0-387-84858-7

[45] Elahifar B, and Hosseini E. A New Approach for Real-Time Prediction of Stick–Slip Vibrations Enhancement Using Model Agnostic and Supervised Machine Learning: A Case Study of Norwegian Continental Shelf. J Pet Explor Prod Technol, 2024; 14: 175-201. https://doi.org/10.1007/s13202-023-01691-6

To whom correspondence should be addressed: Mostafa Bayan, Petroleum Engineering Department, Faculty of Petroleum and Mining Engineering, Suez University, Suez, Egypt, E-mail: Mostafa.BMHe@pme.suezuni.edu.eg