Journal of Engineering Research and Reports

Volume 26, Issue 4, Page 165-176, 2024; Article no.JERR.114132
ISSN: 2582-2926

Predicting Onset of Sand Production in Oil Wells using Machine Learning

Gorei Nkela Ngochindo [a] and Amieibibama Joseph [b*]

[a] *Institute of Petroleum and Energy Studies, University of Port Harcourt, Nigeria.*
[b] *Department of Petroleum and Gas Engineering, University of Port Harcourt, Nigeria.*

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JERR/2024/v26i41123

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/114132

Original Research Article

Received: 02/01/2024
Accepted: 07/03/2024
Published: 27/03/2024

ABSTRACT

Sand production in oil wells is a significant challenge that negatively impacts productivity and compromise equipment integrity. This study explores the application of Optimized Support Vector Machine (SVM) binary classification algorithm to predict the onset of sand production in oil wells. A dataset from 63 oil wells was utilized, and class labels were determined based on the bulk and shear modulus product. The model development incorporated geological and mechanical parameters that could influence sand detachment such as: Young's modulus, Poisson's ratio, minimum and maximum horizontal stresses, overburden pressure, pore pressure, depth, fracture gradient, and formation strength. Instances above the threshold of 8E+11 were classified as indicative of no sand production, while those below were considered potential sand production scenarios. The SVM model demonstrated remarkable accuracy in predicting sand production onset, trained and tested rigorously with field data. The model's accuracy was evaluated using statistical parameters, such as: accuracy (ACC), sensitivity (SE), specificity (SP), and Matthew's Correlation Coefficient (MCC). From the results, the model achieved a score of 1 across all parameters, indicating high reliability and accuracy in sand production prediction. The practical implications of this model are significant, offering assistance to completion engineers in making proactive decisions regarding sand control strategies. Furthermore, the integration

**Corresponding author: Email: amieibibama.joseph@uniport.edu.ng;*

of this model into oil and gas industry processes can optimize operational efficiency by foreseeing potential sand production events, hence, preventing production impairment and ensuring loss prevention.

Keywords: Sand production; machine learning; SVM, bulk and shear modulus product (GKb); classification; young modulus.

ABBREVIATIONS

ACC	:	*Accuracy*
COH	:	*Cohesive strength of the formation (Kg/cm^2)*
CTD	:	*Critical Total Drawdown (Kg/cm^2)*
D	:	*Depth (ft)*
E	:	*Young's modulus*
FS	:	*Formation strength (psi)*
EOHS	:	*Effective overburden horizontal stress (Kg/cm^2)*
G	:	*Shear modulus*
Kb	:	*Bulk modulus*
Max HS	:	*Maximum horizontal stress (psi/ft)*
Min HS	:	*Minimum horizontal stress (psi/ft)*
MCC	:	*Matthew's correlation coefficient*
OP	:	*Overburden pressure (psi/ft)*
PP	:	*Pore pressure (psi/ft)*
SE	:	*Sensitivity*
SP	:	*Specificity*
SVM	:	*Support vector machine*
TN	:	*True negative*
TP	:	*True positive*
TT	:	*Transit time (microsecond/ft)*
TVD	:	*True vertical depth (feet)*

Greek Letter

v	:	*Poisson's ratio*

1. INTRODUCTION

Sand production in oil wells is a long-standing challenge for the oil and gas industry, significantly reducing wellbore productivity [1]. It occurs when sand from the reservoir is transported along with hydrocarbons to the wellbore, where it can accumulate and cause operational problems. Predicting the onset of sand production is critical for maintaining and optimizing well operations. The prediction of sand production from gas and oil wells has been studied using analytical, empirical, and numerical methods. Numerical techniques utilize the finite element method (FEM) or finite difference method to analyze three-dimensional stresses and material behaviour like plasticity, and fluid flow. This method, however, is time-consuming and highly complex [2].

The most thorough and exact-solution approaches are the analytical methods; yet they are difficult to apply to complicated problems [3]. One of the drawbacks of this method in sand production prediction is that it ignores the variation of stress in different directions in a material and assumes symmetrical geometry as boundary conditions [4]. Consequently, this method may fail to describe the sanding risk associated with the orientation of the borehole if it ignores the fundamental impact of stress anisotropy on sanding. Regardless of the intricacies of these models, assumptions or approximations in the absence of real data generally make them unreliable and erroneous [5]. The prediction of sand production using empirical methods is based on field observations and well data. The correlation between sand production, well data, field and operation parameters is established using sand prediction algorithms that rely on field experiences and is categorized into: correlations with one, two, and multiple parameters [6].

The literature contains a few models and correlations that can be used to forecast the critical total drawdown (CTD), which is a measure of when sand production will start. The CTD is the maximum difference that the formation can withstand without producing sand. It is the difference between the minimum well bottom hole flowing pressure and reservoir pressure. To forecast the CTD, some studies employed analytical models such as Mohr-Coulomb and modified Lade; however, the models contain certain assumptions, such as the mechanical properties of the rock formation are homogeneous and isotropic, making them lack accuracy [7,8]. Other models such as artificial neural networks (ANNs), feed-forward backpropagation networks (BPN), generalized regression neural networks, multiple linear regression (MLR), and the genetic algorithm MLP (GA- MLR) have been applied to predict CTD, but these models are proven to lack accuracy [4].

Some studies have used the fuzzy logic (FL) approach in predicting sand production [4]; and this approach has been shown to have higher accuracy and reliability. However, FL has some limitations in capturing the nonlinear and stochastic behaviour of sand production, as well as difficulties in explaining the fuzzy logic results and integrating the fuzzy logic with other methods or tools [9].

The emergence of machine learning and artificial intelligence has opened new possibilities for predicting sand production with greater accuracy. Machine learning (ML) algorithms, such as regression, classification, and neural networks, can analyse vast datasets encompassing geological, operational, and production data to identify hidden patterns and relationships. By leveraging this technology, it is possible to develop predictive models that can offer real-time sand production onset forecasts, allowing for proactive intervention to prevent damage and optimize well performance.

ML approaches have found growing applications in petroleum engineering, including in the area of sand production prediction. For example, a recent study by Alakbari et al.[10], combined the response surface methodology (RSM) and support vector machine (SVM) to develop a more accurate prediction of CTD in gas wells, considering four parameters: total vertical depth (TVD), transit time (TT), cohesive strength (COH), and effective overburden vertical stress (EOVS). The model was shown to be more accurate than existing models in the literature.

Efforts have been made by researchers to predict sand production using machine learning binary classification algorithms. Ngwashi et al. [11] introduced a two-layered Artificial Neural Network (ANN) employing a back-propagation algorithm, implemented in the PYTHON programming language. The model utilizes 11 geological and reservoir parameters associated with the onset of sanding, encompassing depth, overburden, pore pressure, maximum and minimum horizontal stresses, well azimuth, well inclination, Poisson's ratio, Young's Modulus, friction angle, and shale content. Another study by Belyadi and Haghighat [12] involved the development of a K-Nearest Neighbour (KNN) binary classification model for predicting sand production onset, utilizing data from 29 wells. The parameters considered were TVD, TT, COH, EOVS, bottom-hole flowing pressure (BHFP), drawdown pressure (DD), gas flow rate (Q_g), shots-per foot (SPF), water flow rate (Q_w) and perforation interval (Hperf). Similarly, Abilov et al.[13], applied a machine learning algorithm, specifically a Random Forest Classifier, for the detection of sand production events using parameters from an oil field in the South Caspian Basin. These models were shown to have good performance, with accuracy range between 70% to 80%.

.

The desire for better accuracy has driven the popularity of SVM models in binary classification, which has seen their application in multiple fields [14]. In the area of sand production control and management, researchers have applied SVM in sand prediction. For example, Olatunji and Michael [15] utilized the SVM algorithm to forecast sand production in oil and gas reservoirs, establishing its robust classification capabilities with 100% accuracy. In contrast to the utilization of geomechanical properties employed in this study, their investigation incorporated factors like flow rate, volume of sand produced per foot, and perforation parameters. Similarly, Gharagheizi et al. [16] applied the SVM model to predict the onset of sand production in petroleum reservoirs, attesting to its high classification effectiveness with a 100% accuracy score. Their study encompassed variables such as bottom hole flowing pressure, flow rates, and the volume of sand produced per foot (shots per foot).

Hence, in this study a new robust and more accurate model for predicting the onset of sand production using bulk and shear modulus product (GKb) is developed. The model was developed using parameters such as Young’s modulus, Poisson’s ratio, minimum and maximum horizontal stresses, overburden pressure, pore pressure, depth, fracture gradient, and formation strength.

2. METHODOLOGY

A data set that comprised 63 wells from Niger Delta Basin was used for the modelling. Niger Delta is a hydrocarbon-rich basin in Nigeria, developed in the early cretaceous geologic period [17]. The reservoir deposits are composed of shoreface and channel sandstone facies [18]. The Basin has been explored and exploited for hydrocarbon for over six decades [19].

This work incorporates a quantitative approach using machine learning techniques to predict sand production onset in oil wells. It was carried out following the steps shown in Fig. 1. This was achieved using the Classification Learner application of MATLAB R2023b. The datasets were obtained and prepared. The preparation included the definition of predictors (explanatory variables) and target variables. The dataset were further subdivided into training and testing datasets, and then normalized to improve the performance and training stability of the model. The SVM model was developed to predict the onset of sand production based on an 8.0E +11 psi^2 threshold [20], obtained from the product of bulk modulus and shear modulus (GKb). After the development and validation of the model, the model was then tested using field data.

The validity and reliability of the machine learning model were assessed through testing and validation against field datasets, cross-validation techniques and evaluation using statistical metrics to ensure robustness, reliability and accuracy. Data analysis encompassed the utilization of machine learning algorithms to predict sand production onset based on reservoir parameters. Descriptive and inferential statistics were also employed for data analysis. To choose the model used in this study, several other models were tried and ranked using the training dataset. The model with the highest accuracy was then chosen, and in this case, it was the optimized support vector machine (SVM). However, it is worth mentioning that limited availability of comprehensive datasets, especially in capturing dynamic real-time operational variables or geological intricacies, could impact negatively on the predictive accuracy of machine learning algorithms.

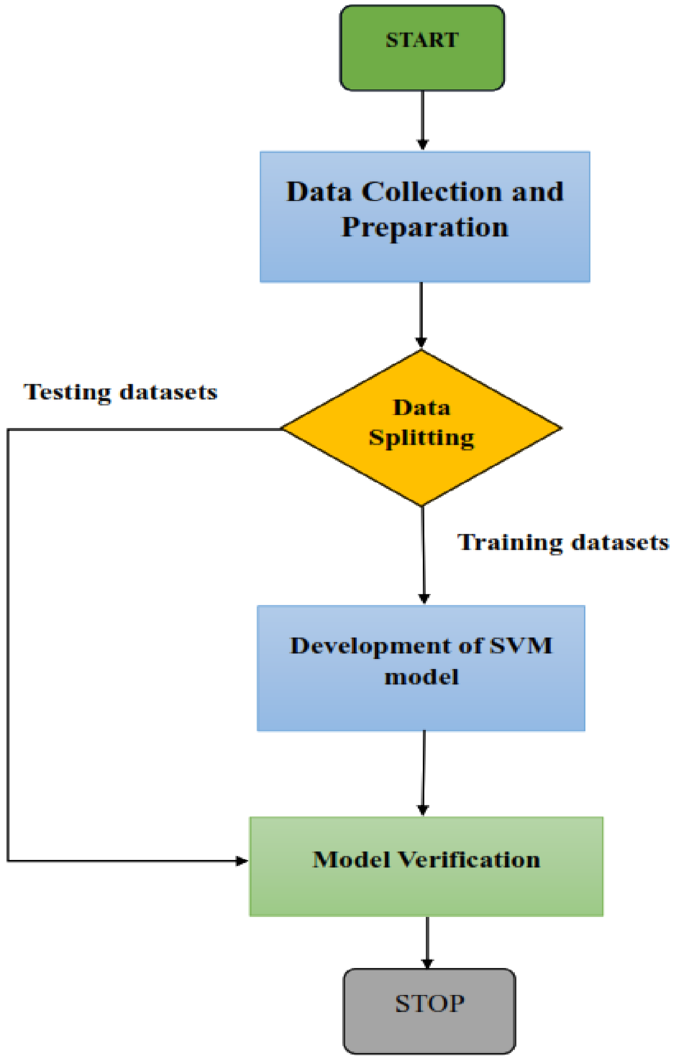

Fig. 1. Flow chart showing the methodology adopted in this study

3. RESULTS AND DISCUSSION

A descriptive statistic of the input parameters used in the development and testing of the SVM model is presented in Tables 1 and 2. The parameters are depth (D), Poisson's ratio (ν), Young's modulus (E), minimum horizontal stress (Min HS), maximum horizontal stress (Max HS), overburden pressure (OP), pore pressure (PP), fracture gradient (Frac. Grad), and formation strength (FS). For the Testing datasets, the depth, pore pressure, minimum and maximum horizontal stresses, and formation strength do not have repetitive values, hence no mode values. Table 3 shows a summary of the specifications of the developed SVM model. One-vs-one cross-validation is employed in developing the model because the dataset is imbalanced with more instances of sand production observations than no sand production. Imbalanced datasets are common in various real-world applications [21].

The correlation coefficient (R) is calculated to evaluate and determine the importance of each input parameter in predicting the output. Classes 1 and 0 are used to represent sand production and its absence, respectively. They are labeled based on the GKb product values, with a threshold of $8.0\mathrm{E}+11\mathrm{psi}^{2.}$ If GKb is greater than the threshold, there is no potential for sand production. If GKb is less than the threshold, there is a potential for sand production. The GKb is a strong function of Young's modulus, formation strength, and depth, where the correlation coefficients were 0.905, 0.805, and 0.796, respectively. In addition, it is a function in the opposite direction of Poisson's ratio, where the R was 0.522, as shown in Fig. 2. A negative R implies that the variables are inversely related [22].

A scatter plot of Young's modulus (E) against GKb is shown in Fig. 3. Some data points exhibit a linear relationship and others appear to be widely spread across the graph area. To evaluate the performance of the model, the confusion matrix was plotted as shown in Fig. 4, and the following parameters were obtained: true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), true positive rate (TPR) and false negative rate (FNR). The true positive rate (TPR) and false negative rate (FNR) denote the percentages at which the model forecasts true positives (TP) and false negatives (FN), respectively. The True Positive Rate (also known as Sensitivity or Recall – REC) is determined by dividing the number of correct positive predictions (TP) by the total count of positives (P). The highest achievable TPR is 1.0, while the lowest is 0.0. The True Negative Rate or Specificity (TNR), is computed by dividing the accurate negative predictions (TN) by the total number of negatives (N). The optimal specificity is 1.0, while the least favourable is 0.0.

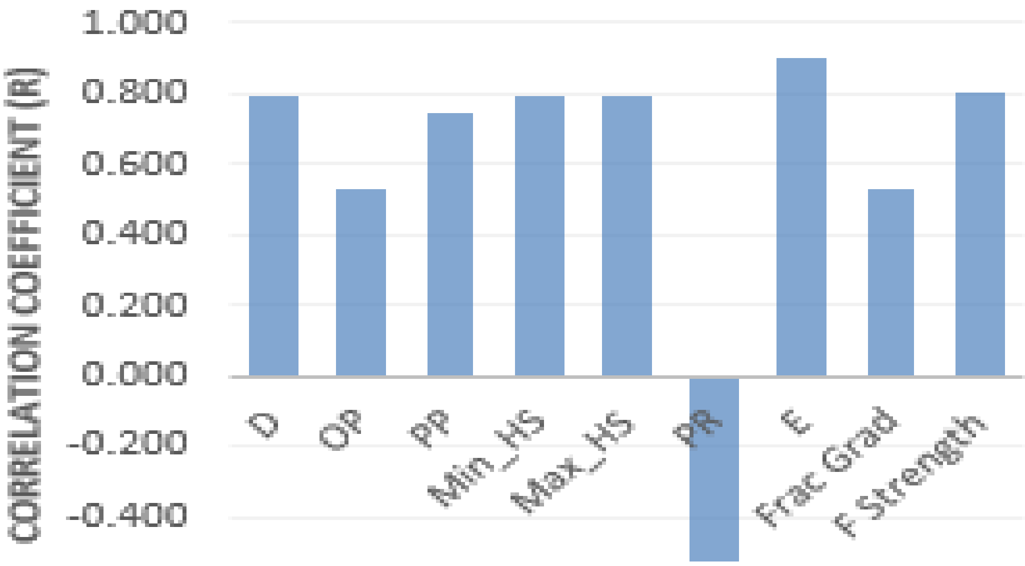

Fig. 2. Relative importance of input parameters to the model's output

Table 1. Statistical analysis of the data used in developing the svm model

Parameters	D (ft)	OP psi/ft	PP psi/ft	Min HS psi/ft	Max HS psi/ft	v	EM psi	Frac. Grad (psi/ft	FS psi
Minimum	4000	0.8211	0.397	0.6297	0.6797	0.204	0.107	0.304	2256
Maximum	14875	0.982	0.477	0.9159	0.96	0.28	4.0414	0.76	9164
Mean	10144.4	0.9041	0.447	0.835	0.888	0.256	0.487	0.576	6104.2
Median	10232	0.902	0.439	0.88	0.93	0.25	0.456	0.6	6085
Mode	5243	0.928	0.439	0.89	0.94	0.25	0.702	0.62	8708.4
Range	10875	0.1609	0.08	0.2862	0.2803	0.076	3.9344	0.456	6908
STD	3366.8	0.053	0.020	0.065	0.068	0.016	0.535	0.093	2105.5

Table 2. Statistical analysis of the data used in testing the model

Parameters	D (ft)	OP psi/ft	PP psi/ft	Min HS psi/ft	Max HS psi/ft	v	EM psi	Frac. Grad (psi/ft)	FS (psi)
Minimum	5843	0.821	0.369	0.614	0.664	0.156	0.107	0.304	3548.0
Maximum	14500	0.985	0.471	0.922	0.96	0.291	4.310	0.820	9423.0
Mean	9177.4	0.916	0.418	0.777	0.825	0.268	0.990	0.597	5576.8
Median	7529.5	0.932	0.412	0.779	0.829	0.28	0.278	0.595	4480.0
Mode		0.981				0.28	0.115	0.590	
Range	8657	0.164	0.102	0.308	0.296	0.135	4.2026	0.516	5875.0
STD	3543.7	0.068	0.037	0.111	0.109	0.040	1.572	0.127	2357.7

Table 3. Specifications of the SVM model

Parameters	Description/Value
Model	Optimizable Support Vector Machine (SVM)
Input	9
Output	1
Multiclass Coding	One-vs-one cross-validation
Iterations	30
Optimizer	Bayesian Optimization

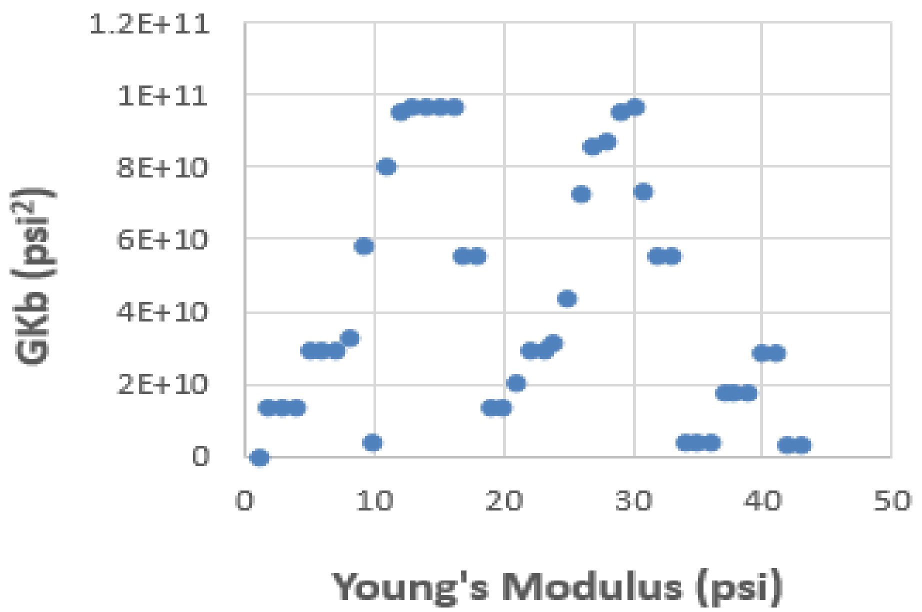

Fig. 3. Scatter plot of young's modulus against

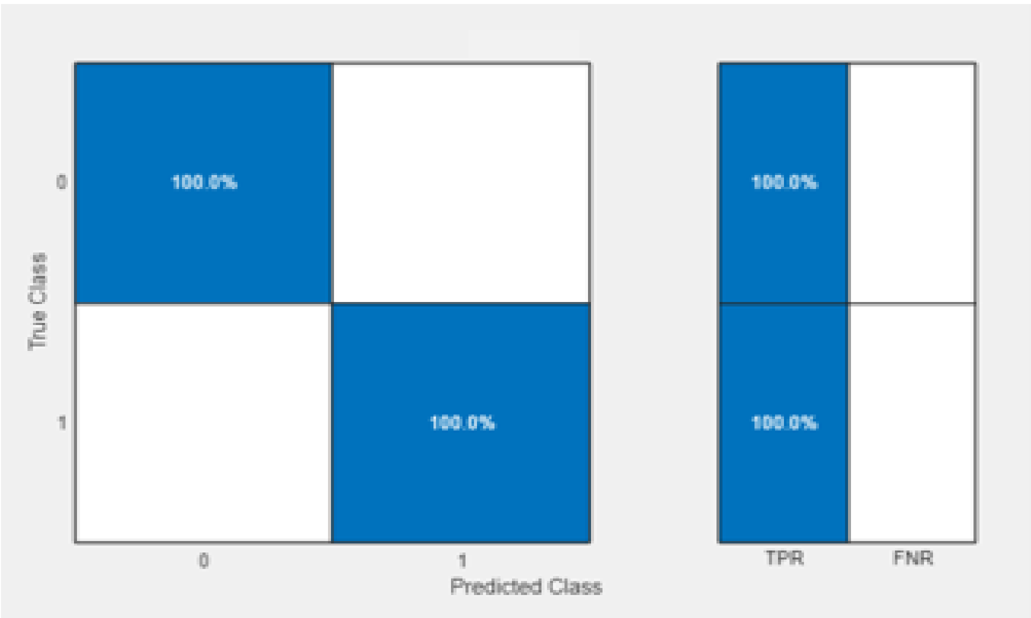

Fig. 4. Confusion matrix showing the performance shear and bulk modulus product of the SVM model in predicting sand production and the absence of sand production for the test dataset

Table 4 displays the outcomes of the trained SVM model on the test dataset while Table 5 is a summary of the results from the statistical parameters used in evaluating the validity and reliability of the developed model. Within the test dataset, there exist ten instances where seven wells showed sand production (labelled as 1) and three wells showed no sand production (labelled as 0). When the model correctly predicts a value of 1 or sand production, it is termed a True Positive (TP). In this scenario, there are six instances of TP. Likewise, when the model accurately predicts values of 0 or no sand production, it is referred to as True Negative TN. There are three TNs. Should the model wrongly predict a value of 1 or sand production, it is identified as False Positive (FP) or a type I error. Additionally, if the model erroneously predicts a value of 0 or no sand production, it's termed a False Negative (FN) or a type II error. In this instance, there are six TP, one TN, and two FP. Notably, there were no instances of FN in the model's predictions.

Table 4. Model predictions for test dataset comprising seven sand-producing wells and three non-sand-producing wells

Y_ Test	Y_Test-Predict	Prediction Status
0	0	TN
1	1	TP
0	0	TN
1	1	TP
1	1	TP
1	1	TP
0	0	TN
1	1	TP
1	1	TP
1	1	TP

The Receiver Operating Characteristic (ROC) curve, as shown in Figure 5, is used to measure the performance of the model and thus, validate it. This also provides information on the accuracy of the binary classification. Accuracy measures the ratio of accurate predictions to the total number of predictions made.

The Receiver Operating Characteristic (ROC) employs several parameters, including goodness-of-fit, robustness, and predictive capability in classifying data. Goodness of fit is assessed through measures such as accuracy (ACC), sensitivity (SE), and specificity (SP). Accuracy (ACC) serves as a straightforward indicator of the quality of the Support Vector Machine (SVM), representing the ratio of correctly assigned cases. Sensitivity (SE) quantifies the model's accuracy in correctly classifying observed instances, while specificity (SP) measures the proportion of correctly classified cases relative to the entire population. The computation of ACC, SE, and SP are performed using the following equations [23,24]:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP}} \tag{1}$$

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{3}$$

Another binary classification evaluation metric used in measuring the quality of the classification model developed in this work is the Matthew's correlation coefficient (MCC) [23,24]:

$$\mathrm{MCC} = \frac{(\mathrm{TP} \times \mathrm{TN}) - (\mathrm{FP} \times \mathrm{FN})}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \tag{4}$$

The computed values of ACC, SE, SP and MCC are shown in Table 5.

Table 5. Statistical classification parameters obtained for developed SVM Model

Statistical Parameters	Training Set	Test Set
TP	44	7
TN	9	3
FP	0	0
FN	0	0
SE	1	1
SP	1	1
ACC	1	1
MCC	1	1

4. DISCUSSION

The findings of this study shows the relationship between the geomechanical parameters – Young's modulus, Poisson's ratio, minimum and maximum horizontal stresses, depth, overburden pressure, pore pressure, fraction gradient and formation strength and sand production in oil wells. This agrees with works of Ajayi et al [25] and Lawson-Jack [26] that shows the effectiveness of using geomechanical properties to predict sand production, and therefore, highlights their crucial role in determining the onset of this phenomenon. The pronounced correlation observed for Young's modulus (Fig. 2) is primarily due to its unique position at the intersection of stress and strain, both geomechanical properties that significantly influence sand production.

As shown in Fig. 3, the nonlinear relationship of the data points suggests a complex relationship between the predicting variable, in this case, Young's modulus, and sand production. The sand production phenomenon has been described as a complex process [27]. It was observed that the other predictors followed a similar trend. As indicated in Table 5, the values obtained for FP, FN, SE, SP, ACC, and MCC in both the training and testing data sets are identical. This consistency shows that overfitting did not occur during the training stage. Additionally, as presented in Table 4, none of the instances in the original field dataset were incorrectly classified, suggesting that the model effectively discriminates between classes with minimal errors and exhibits a well-balanced trade-off between precision and recall, showcasing strong predictive power.

Upon examining the data shown in Tables 4 and 5; and Fig. 3 and 4, it can be seen that the developed SVM model demonstrates considerable promise in forecasting sand production. Consequently, the integration of SVM modelling is a valuable approach, assisting completion Engineers in devising timely sand control strategies while minimizing production disruptions. This, in turn, holds the promise of enhancing operational efficiency and optimizing the management of sand-related challenges within the industry. The model developed in this study can be compared with three binary classification models from previous studies[11,12,13]. The performance of these models is compared with the SVM model developed in the present study, particularly focusing on accuracy metrics, as shown in Table 6.

Table 6. Comparison between the Proposed SVM model and other models

MODEL	ML ALGORITHM	ACCURACY (%)
Abilov et al. [13]	Random Forest Classifier (RFC)	77
Belyadi and Haghighat [12]	K-Nearest Neighbour (KNN)	78
Ngwashi et al. [11]	Artificial Neural Network (ANN)	80
This study	Support Vector Machine (SVM)	100

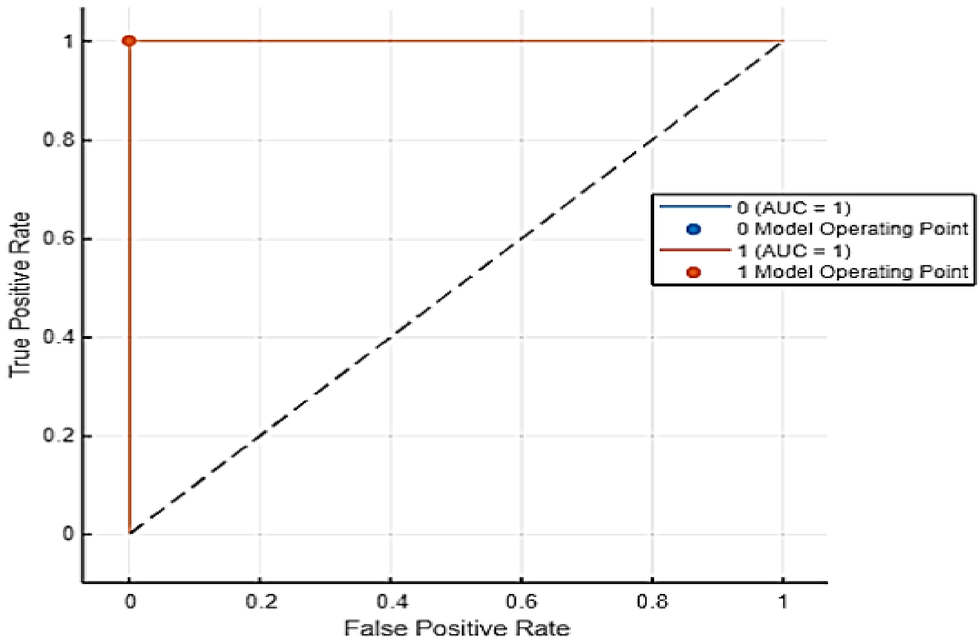

Fig. 5. Receiver Operating Characteristic (ROC) curve for the developed optimized SVM model based on the test datasets

The SVM model developed in this study outperformed the other three models, achieving an accuracy of 100% compared to the next-best model, the ANN, with an accuracy of 80%. The accuracy of the KNN and RFC models are 78% and 77% respectively. The SVM model has shown to be a powerful tool in the prediction of sanding in oil and gas wells. The high accuracy achieved by SVM is due to its effectiveness with binary classification achieved by finding the optimal hyperplane (the best-fitting dividing line) that separates the data into two categories [14]. SVM can also handle high-dimensional data, and non-linear classification problems using kernel functions, and performs well with small data samples [28]. These are in agreement with the study by [11] which asserts that SVM is a better machine-learning tool for binary classification when compared with other machine-learning tools used for predicting sand production. The performance of the SVM model underscores the robustness and reliability of machine-learning techniques in forecasting sand production in oil wells. Such accuracy not only solidifies the efficacy of the SVM model but also signifies its potential practical application within the oil industry.

5. CONCLUSION

The application of the Support Vector Machine (SVM) algorithm in combination with the product of bulk and shear modulus (GKb) as the criteria for sand production prediction has yielded promising results. Applying the binary classification methodology used in machine learning, and through a robust analysis of historical data on well conditions and instances of sand production, the SVM model has shown a high degree of accuracy in anticipating potential sand production. By leveraging the relationship between the bulk and shear modulus, the model successfully identified patterns and correlations, enabling proactive identification of potential sand production occurrences. The study demonstrated the SVM model's ability to effectively differentiate between conditions for potential sand production and those where the risk remains minimal. The performance of the model was validated through rigorous testing with statistical metrics for evaluating classification machine learning models which include accuracy, sensitivity, specificity, and Mathew's correlation coefficient. These metrics showed good agreement in establishing the reliability and accuracy of the Model which was also validated with field datasets, indicating its reliability in real-world scenarios.

The study underscores the significance of leveraging machine learning, specifically the SVM model based on the product of bulk and shear modulus, in the oil industry's quest to preemptively address sand production issues in oil wells. By accurately predicting potential occurrences of sand production, this approach equips operators with the foresight necessary to implement timely preventive measures, thereby minimizing operational disruptions and associated costs. However, incorporating real-time measurements and predictions by integrating all known variables that could influence the disengagement of sand particles during production could help in addressing sand management issues in oil wells.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Salahi A, Dehghan AN, Sheikhzakariaee SJ, Davarpanah A. Sand production control mechanisms during oilwell production and construction. Petroleum Research. 2021;6(4):361-367.
2. Shahsavari MH, Khamehchi E, Fattahpour V, Molladavoodi H. Investigation of sand production prediction shortcomings in terms of numerical uncertainties and experimental simplifications. J. Petroleum Sci. Eng. 2021;207:109147.
   DOI: 10.1016/j.petrol.2021.109147
3. Creosteanu A, Gavrila G, Creosteanu L. Comparison between an analytical method and two numerical methods on a given electrostatic potential determination problem. In 2012 15th International Symposium on Antenna Technology and Applied Electromagnetics. Toulouse, France. 2012;1-6.
   DOI: 10.1109/ANTEM.2012.6262310
4. Alakbari FS, Mohyaldinn ME, Ayoub MA, Muhsan AS, Hussein I A. A robust fuzzy logic-based model for predicting the critical total drawdown in sand production in oil and gas wells. PloS one. 2021;16(4):e0250466.
5. Abbas AK, Baker HA, Flori RE, Al-hafadhi H, Al-haideriN. Practical approach for sand-production prediction during production. In ARMA US Rock Mechanics/Geomechanics Symposium ARMA; 2019.

6. Khamehchi E, Kivi IR, Akbari M. A novel approach to sand production prediction using artificial intelligence. Journal of Petroleum Science and Engineering. 2014;123:147-154.
7. Moreno FG, Guizada P, Abdul Aziz AA, Khanferi RN. Application of critical drawdown pressure prediction in completion design to minimize sanding in a clastic gas reservoir in Saudi Arabia. In SPE Kingdom of Saudi Arabia Annual technical symposium and exhibition. Society of Petroleum Engineers; 2017.
8. Zhang R, Shi X, Zhu R, Zhang C, Fang M, Bo K., et al.. Critical drawdown pressure of sanding onset for offshore depleted and water cut gas reservoirs: modelling and application. Journal of Natural Gas Science and Engineering. 2016;34:159–169.
9. Subbiah S.K, Samsuri A, Mohamad-Hussein A, Jaafar MZ, Chen YR, Kumar RR, Root cause of sand production and methodologies for prediction. Petroleum 2021;7(3):263-271.
10. Alakbari FS, Mohyaldinn ME, Ayoub MA, Muhsan AS, Abdulkadir SJ, Hussein IA, Salih AA. Prediction of critical total drawdown in sand production from gas wells: Machine learning approach. The Canadian Journal of Chemical Engineering. 2023;101(5):2493-2509.
11. Ngwashi AR, Ogbe DO, Udebhulu DO. Evaluation of machine-learning tools for predicting sand production. In SPE Nigeria Annual International Conference and Exhibition (p. D031S016R001); 2021
12. Belyadi H, Haghighat A. Machine learning guide for oil and gas using python: a step-by-step breakdown with data, algorithms, codes, and applications. Gulf Professional Publishing; 2021.
13. Abilov E, Rzayeva G, Alkhasli S, Pashayev A, Allahverdiyev P. Real-time detection of sand production event by application of machine learning techniques. European Association of Geoscientists & Engineers, In 84th EAGE Annual Conference & Exhibition 2023;(2023):1-5.
14. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing. 2020;408: 189-21 5.
15. Olatunji OO, Micheal O. Prediction of sand production from oil and gas reservoirs in the Niger Delta using support vector machines SVMs: A binary classification approach. In SPE Nigeria Annual International Conference and Exhibition, D023S026R001; 2017
16. Gharagheizi F, Mohammadi AH, Arabloo M, Shokrollahi A. Prediction of sand production onset in petroleum reservoirs using a reliable classification approach. Petroleum. 2017;3(2):280-285.
17. Deng RJ, Deng YH, Yu S, Hou DJ. Hydrocarbon geology and reservoir formation characteristics of Niger Delta Basin. Petroleum Exploration and Development. 2008;35(6):755-762.
18. Ogbe OB. Sequence stratigraphic controls on reservoir characterization and architectural analysis: A case study of Tovo field, coastal swamp depobelt, Niger Delta Basin, Nigeria. Marine and Petroleum Geology. 2020;121:104579.
19. Adedapo J, Ikpokonte A, Schoeneich K, Kurowska E. An estimate of oil window in Nigeria Niger Delta Basin from recent studies. Gas. 2014;5(11):12-18,
20. Khamehchi E, Reisi E. Sand production prediction using ratio of shear modulus to bulk compressibility (case study). Egyptian Journal of Petroleum. 2015;24:113-118.
21. Cateni S, Colla V, Vannucci M. A method for resampling imbalanced datasets in binary classification tasks for real-world problems. Neurocomputing. 2014;135:3 2-41.
22. Sauro J, Lewis JR. Quantifying the user experience: Practical statistics for user research, 2nd Edition, Elsevier. Morgan Kaufman; 2016.
23. Kuhn M, Johnson K. Measuring Performance in Classification Models. In: Applied Predictive Modeling. Springer, New York, NY; 2013. Available:https://doi.org/10.1007/978-1-4614-6849-3_11
24. Vujović Ž. Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications. 2021; 12(6):599-606.
25. Ajayi TJ, Adeogun OY, Omeru T, Predicting sanding potential using empirical method in Ebendo Field, Niger Delta, Nigeria. Journal of Applied Science and Environmental Management. 2022; 26(9):1557-1564.

26. Lawson-Jack O, Uko DE, Tamunobereton-Ari I, Alabraba MA. Geomechanical characterization of a reservoir in part of Niger Delta, Nigeria. Asian Journal of Applied Science and Technology. 2019; 3(1):10-30
27. Ubuara DO, Olayinka YA, Emujakporue GO, Soronnadi-Ononiwu G.C, Evaluation of formation susceptibility and sand production potential in an offshore field, Niger Delta Basin, Nigeria. Energy Geoscience. 2024;5(1):100213.
28. Chauhan VK, Dahiya K, Sharma A. Problem formulations and solvers in linear SVM: A review. Artificial Intelligence Review. 2019;52(2):803-855.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/114132