Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **3**, September **2025**

Develop Oil and Gas Production Forecasting Models based on Deep Learning and Machine Learning

Raed Majeed

Department of Computer Information Systems
University of Sumer
Dhi-Qar, Iraq
raed.m.muttasher@gmail.com

Hiyam Hatem

Department of Computer Science
University of Sumer
Dhi-Qar, Iraq
hiamhatem2005@gmail.com

Mohammad Kaisb

Department of Computer Science
University of Sumer
Dhi-Qar, Iraq
mohammad.kaisb@uos.edu.iq

Received May 22, 2025, revised July 6, 2025, accepted July 9, 2025.

Abstract. *Energy companies need to forecast oil and gas production to increase their output. Forecasting production rates helps petroleum businesses plan operations, optimise production, and allocate resources. Researchers have used numerical reservoir simulation (NRS) and decline curve analysis (DCA) to predict oil and gas output. These algorithms faced obstacles such a time-consuming and precise static model, many dynamic model parameters, and unknown correctness. Build and assess machine learning and deep learning models to anticipate oil and gas output to avoid these challenges. The suggested system leverages predictive and inferential data analytics to make faster, more accurate decisions. They mean successfully forecasting future events from past data. This paper presents eight methodologies, including four machine learning models: decision tree regressor (DTR), random forest regressor (RFR), k-nearest neighbors (KNN), and extreme gradient boosting (XGBoost). And other four deep learning models: artificial neural networks (ANN), recurrent neural network (RNN), long short-term memory (LSTM) and 1-D CNN-based Regressor investigated. The goal is to construct a model that outperforms DCA and NRS for faster and more accurate oil and gas output insights. The best model choice may differ based on the distinct characteristics of the data being examined. Hence, conducting experiments with several models and evaluating their efficacy is crucial to determine the most appropriate choice. The results indicated a superiority of all models along with lower error rates in both Mean Squared Error (MSE) and Mean Absolute Error (MAE). the 1-D CNN model achieved the highest accuracy in R^2 (Oil) of 0.9951, while the KNN model performed the best for R^2 (Gas) with a 0.9988.*
Keywords: Deep Learning (DL), Machine Learning (ML), Oil and Gas Production, Random Forest , 1-D CNN-based Regressor, K-Nearest Neighbors (KNN).

1. **Introduction.** Oil is the main fuel and essential for many manufacturing processes. Oil and its derivatives have substantially affected global energy supplies. This trend is expected to continue, increasing oil consumption relative to alternative energy sources [1]. Oil production begins with site discovery, continues through extraction, and ends with product distribution to businesses and the public [2]. Oil and gas must be extracted and utilized for energy through many procedures. The oil and gas business has three basic production processes: Upstream industry, Midstream and The downstream.
Prediction is essential for field development planning because it offers production data for facility capacity design, drilling timetables, and economic assessments. Past active and no active well production data is in high demand for production estimates [3]. Governments and organizations need production forecasts to create economic policies [4]. Oil and gas output forecasts involve complex reservoir numerical simulations and engineering studies [5]. Precision prediction is a major task for oil reservoir monitoring and improvement. Traditional petroleum industry methods include numerical reservoir simulation (NRS) and decline curve analysis (DCA) [6]. Due to field vastness and reservoir complexity, these models took longer to make judgments [7]. For years, NRS has been used. NRS models are hard, time-consuming, and need a precise static model and many dynamic model parameters [8]. DCA predicts oil and gas production traditionally [9]. Fitting the simulated cumulative production rate yielded DCA model parameters [10].

Based on previous production, DCA predicts well or field production. Predictions may be used to evaluate the economics of future output and help choices to abandon a well or field by better using computing resources. Machine and deep learning are changing the oil industry. AI lets computers analyses and decides. Focusing on ML for forecasting, sophisticated ML and DL algorithms, and large data collection from various industrial instruments has a promising future in solving oil and gas sector problems [11]. Machine Learning and Deep learning has significantly advanced the fields of image analysis and computer vision in the development of highly accurate and efficient image recognition and classification models [12]. In recent years, oil and gas researchers have used ML and DL algorithms for quick evaluation and output forecasts [13]. Predictive and inferential data analytics in ML and DL enable faster, more accurate decisions. The oil and gas business is quickly integrating data analytics to enhance decision-making [14].

The general structured of this research can be summerized as: Section 2 provides the most resent related work on the research filed Oil and Gas Forecasting. Section 3 defines the dataset employed for the research identifying data type and their features. Section 4 presents the proposed methodology, including the data pre-processing, all of the AI models involved, the performance metric used. Section 4 presents the experimental results, the comparisons with baseline models. Section 5 gives a discussion. Section 6 concludes the paper and refering to the potential directions of future works.

2. **Related Works.** C. Tan et al. (2021) employed the ML techniques random forest (RF), back propagation (BP) neural network, support vector regression (SVR), extreme gradient boosting (XGBoost), light gradient boosting machine (Light GBM), and multivariable linear regression. Data from 137 wells in the WY shale gas block in Sichuan, China was used to evaluate and optimise ML algorithms for solutions. Results show that the XGBoost algorithm's production prediction model is the most effective, with R^2 (0.87). The study's limitations in fracturing productivity prediction and optimising shale gas wells include insufficient regional data [15].

G. Hui et al. (2021) use ML to assess Fox Creek, Alberta, shale gas output. The researchers employed four methods: linear regression, neural networks, XGBoost, and decision tree regressor (DTR). With the greatest coefficient of determination of 0.809, the

additional trees technique won. Due to the case study and this research's constraints, it is impossible to say that these are the ideal findings after doing tests with the dataset and examining the outcomes for every strategy [16].

N. M. Ibrahim et al. (2022) attempted to speed up oil and gas output estimates. DTR, SVR, MLR, XGBoost, PLR, RFR, RNN, and ANN were the eight machine learning and deep learning experiments in their study. According to Saudi Aramco's dataset, RNN, XGBoost, and ANN yielded the greatest results, with R^2 values of 0.926, 0.9012, and 0.9627 for oil, gas, and water This study does not address the ethical concerns of utilising ML and DL models in the oil and gas business, such as environmental or labour displacement [17].

Lan Mai-Cao et al. (2022) compared some ML systems for estimating petroleum output. The ability of four deep neural networks to forecast time-series data has been studied: multilayer perceptron, CNN, LSTM, and GRU. Four conventional ML models: RF, SVR, KNN, and GB. Preprocessed historical data from a well in oil field X, Southern Vietnam was used to create eight prediction models for future petroleum output. Classical ML with SVR was shown to be computationally efficient, with high performance metrics and quick computation time [18].

A. E. Al-Aghbari et al. (2022) used TCN, GRU, RNN, and LSTM models. An ensemble DL model using TCN and LSTM predicted oil volume in one step. In constrained computational resource settings, the suggested technique lowered computational complexity and accuracy, making the model viable. And benchmark models outperformed, residual variance decreased. However, the model is limited to conventional reservoirs [19].

W. Liu et al. (2023) principally uses CNPC's oil production information. Logistic regression (LR), decision tree (DT), RF, KNN, XGBoost (XGB), and gradient boosting decision trees (GBDT) as classification models for reservoir identification and ANN and XGB as regression models for production forecasts could benefit from regional data. Using historical data and ensemble methods, XGB outperforms other reservoir identification models in assessment measures. In accuracy and processing speed, XGB outperforms ANN in estimating cumulative oil output in single wells using effective thickness. The ML technique for reservoir identification and production prediction faces data quality, feature selection, model interpretability, computing resources, overfitting, hyperparameter tweaking, and generalisation issues [20].

Wang et al. (2023) investigated various major oil production factors utilising daily oil production data from 62 oil wells over 10 years. Two models were presented utilising polynomial regression and random forests on these two data sets. The quartic polynomial regression and random forest models were chosen for their low prediction error. RF predictions have a lower error margin than polynomial regression [21].

S. Hosseini et al. (2023) presented an LSTM and 1-D convolutional neural network model for Volve oil field time series production forecasting. The LSTM model outperformed the 1-D CNN model. Using data from all wells during training and testing allows models to be applied to additional wells. This article does not examine model generalisability, which is a shortcoming. The LSTM model yields optimal results with R^2 values of 0.97-0.98 [22].

3. **Dataset Descriptions.** The initial and most vital phase of research technique is data collection. The dataset comprises production information from wells in New York State, USA, spanning the years 1967 to 1999 [23]. The dataset was acquired from the open access source (https://catalog.data.gov/dataset/oil-and-gas-summary-production-data-1967-1999). The dataset has 30.1K rows and 20 columns, including production year, operator, producing formation, county, and production date entered, and active oil

TABLE 1. Feature and their data type

Feature	Data Type
Production Year	int64
Production Date Entered	object
Operator	object
County	object
Town	object
Field	object
Producing Formation	object
Active Oil Wells	int64
Disposal Wells	int64
Active Gas Wells	int64
Taxable Gas (Mcf)	int64
Inactive Gas Wells	int64
Inactive Oil Wells	int64
Self-use Well	object
Water Produced (bbl)	int64
Injection Wells	int64
Purchaser Codes	object
Location	object
Gas Produced (Mcf)	int64
Oil Produced (bbl)	int64

wells, among others. Table (1) presents all data attributes and their classifications within the dataset.

4. **The Proposed Models.** The proposed model framework consists of a series of stages. Figure (1) presents an Illustration of the proposed oil and gas production forecasting models.

4.1. **The Pre-processing Stage.** The pre-processing is an essential stage sine the performance of any ML and DL models are highly affected with the quality of the input data, much features extracted mean much accuracy in performance metrics, in many case the data in the data set need more enhancement and improvements during the training or testing, in the presented approach This stage contains three major steps:

4.1.1. *Data Cleaning.* The main mission of this step to ensure the dataset is clean from any mistakes and free of false and misleading information. It's a procedure for checking that the dataset does not contain any inaccurate information and is employed to handle the noise in the data before starting with ML and DL models. There are two steps including:

1. Detect and locate any absent or incomplete data points within the dataset.
2. Determine appropriate action for missing values, filling them with other appropriate values.

Table 2 identifies the number of unclean data in all dataset features, then selects the unique data to clean it. and, ensure the dataset has been cleaned and there are no errors.

4.1.2. *Data Processing.* In this step, each well location feature has been assigned an index. This index will split into two individual sets, X and Y. After a series of experiments, this process helps to achieve high-accuracy results with ML and DL models.

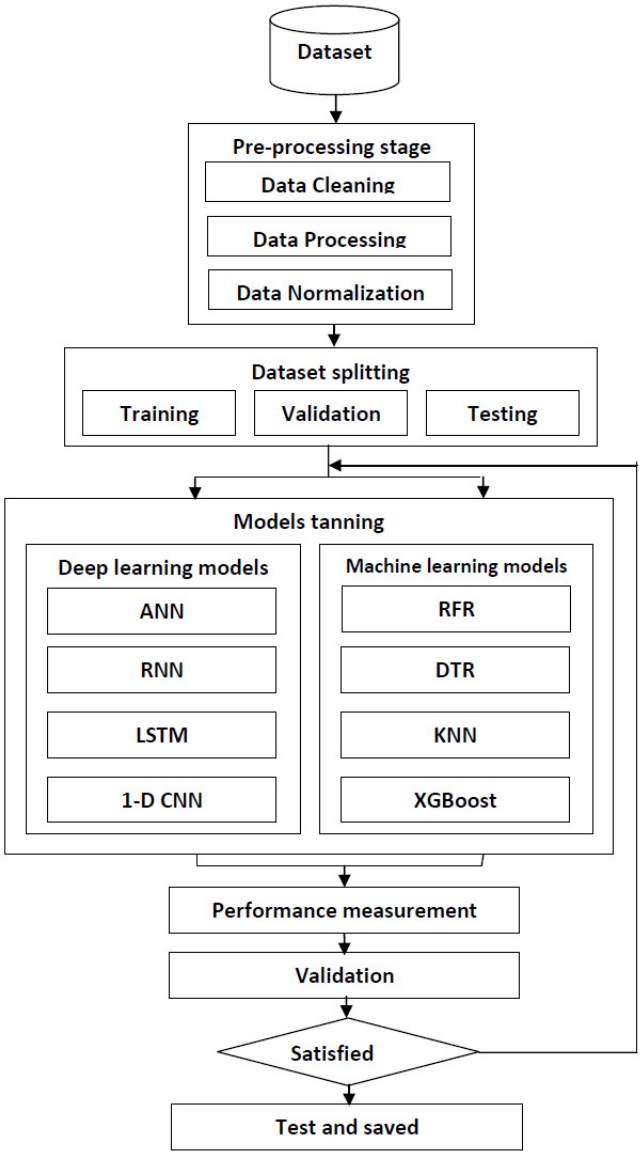

FIGURE 1. Oil and Gas production forecasting models Architecture.

4.1.3. *Data Normalizing.* Data normalization is a technique that enhances the accuracy of ML and DL by adjusting feature values in the dataset to a standard scale. This process also simplifies data analysis and modeling. One common normalization method is L2 scaling, also known as least squares, which converts numerical values into a range between 0 and 1. The data that need normalization have become normalized, like, the features X, Y, active oil well, production year, operator, country, town, and field. Other data that do not require normalization, like injection wells, water produced, and taxable gas, remain unchanged. Figure (2) and figure(3) illustrates the effects of normalization on the first part of the data, showcasing features such as X, Y, active oil well, and production year.

4.1.4. *Splitting Dataset.* The dataset is categorized into three subsets:

1. The training set is the largest set in the dataset (80%) and is used to train presented models and modify the weights by observing and learning the correct output.

TABLE 2. Data Cleaning (DC): (a) the number of unique data before cleaning and (b) the number of unique data after cleaning.

Features before DC	unique data No.	Features after DC	unique data No.
Index	0	Index	0
Production Year	0	Production Year	0
Production Date Entered	0	Production Date Entered	0
Operator	0	Operator	0
County	31	County	0
Town	657	Town	0
Field	1281	Field	0
Producing Formation	660	Producing Formation	0
Active oil wells	0	Active oil wells	0
Inactive oil wells	0	Inactive oil wells	0
Active Gas wells	0	Active Gas wells	0
Inactive Gas wells	0	Inactive Gas wells	0
Injection wells	0	Injection wells	0
Disposal wells	0	Disposal wells	0
Self - use well	619	Self - use well	0
Oil Produced, bbl	0	Oil Produced, bbl	0
Gas produced, Mcf	0	Gas produced, Mcf	0
Water produced, bbl	0	Water produced, bbl	0
Taxable Gas, Mcf	0	Taxable Gas, Mcf	0
Purchaser Codes	11798	Purchaser Codes	0
Location	0	Location	0
dtype: int64		dtype: int64	

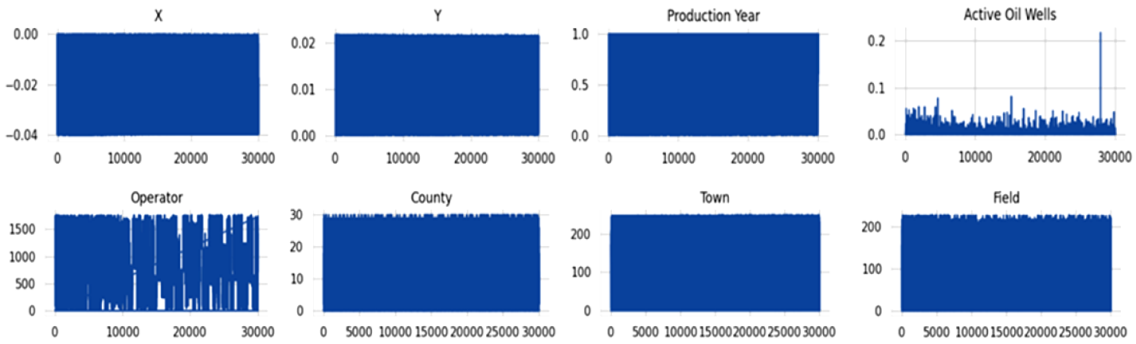

FIGURE 2. Dataset before normalizing.

2. The validation set (10%). This component is employed to evaluate the model by modifying the hyper-parameters and assessing the model's performance throughout the training process.
3. The testing set (10%) is an independent component of the dataset estimates the model's performance on new fresh data.

4.2. **Machine Learning Models.**

4.2.1. *Decision Tree Regressor (DTR).* Both oil and gas production uses the same parameters. Table (3) shows DTR parameter values tested for predictive accuracy. These parameters are critical to model performance. Many tests were run using manually set

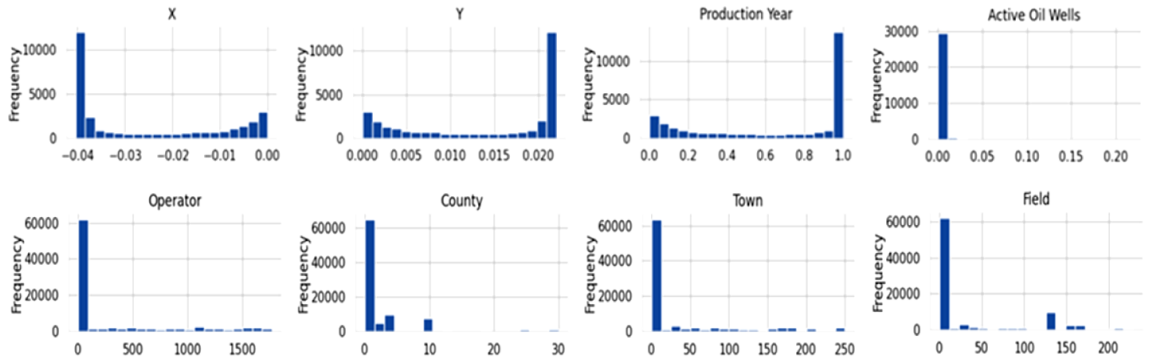

FIGURE 3. Dataset after normalizing.

TABLE 3. Illustrates DTR parameters for Oil and Gas outputs

Parameters	value
Max_depth	500
Random_state	33

TABLE 4. Illustrates RFR parameters for Oil and Gas outputs.

Parameters	value
n_estimators	50
Max_depth	15
Random_state	33

parameters, making parameter selection difficult in this area [24]. Maximum DTR depth is governed by max_depth. A larger number allows for a more sophisticated tree structure with more levels and better outcomes. In this scenario, 500 means the decision tree may go 500 levels deep. The random_state option, which generates random numbers, assures repeatability by fixing the random state. a maximum depth of 500 was selected after conducting grid search optimization combined with cross-validation. This value provided the best trade-off between model accuracy and overfitting, as deeper trees beyond this point did not yield significant improvement in R^2.

4.2.2. *Random Forest Regressor (RFR).* Multiple sub-datasets are used to bootstrap Random Forest Regression predictions. After that, decision trees are generated for each dataset subset [25]. Each sub-decision tree's predictions are combined to create the random forest model's final forecast. Table (4) shows that this model uses generalized error estimates, parameters, and multidimensional data.

1. The n_estimators parameter determines the quantity of trees within the random forest ensemble. Typically, a higher value can improve performance.
2. The max_depth parameter, similar to the counterpart in the decision tree regressor, controls the maximum depth of each decision tree within the forest, established here at 15.
3. The random_state parameter, ensuring result reproducibility by fixing the seed for random number generation, remains constant at 33.

4.2.3. *K -Nearest Neighbors (KNN).* KNN uses a Euclidean distance metric to find fresh data's k nearest neighbors. It then determines the target variable mean (or weighted mean) of these neighboring data points [26]. The mean is used to estimate the new

TABLE 5. Illustrates XGBoost parameters for Oil and Gas outputs

Parameters	value
n_estimator	30
Seed	300

TABLE 6. ANN layers and activation function.

Number of hidden layers	Units of nodes	Activation function
First hidden layer	64	Tanh
Second hidden layer	128	Tanh
Third layer (output)	1	ReLU

data. The same data collection and KNN model are used in oil and gas production. For oil and gas production, the parameters are (n_neighbors = 13, weights = uniform, and algorithm = auto). The prediction neighbor count is determined by n_neighbors. While K symbolises represents the quantity of n_neighbors.

4.2.4. *Extreme Gradient Boosting (XGBoost).* Prior to implementing XGBoost, it is essential to optimise parameters to guarantee optimal model performance. Numerous tests were initially performed with parameters selected manually [27]. Table (5) presents the chosen parameters necessary for optimising oil and gas output. The parameter n_estimators, akin to its equivalent in the random forest regressor, specifies the quantity of trees employed in the boosting procedure. The value is established at 30. The seed parameter establishes the seed for random number generation, thereby ensuring result reproducibility. The value is set to 300 in this scenario.

4.3. **Deep Learning Models.**

4.3.1. *Artificial Neural Network (ANN) Model.* The ANN model comprises two layers and an output layer. The ideal configuration of hidden layers and neurones for the ANN model was established via extensive testing [28]. Table (6) presents the hidden levels of the ANN, the number of nodes in each layer, and the activation function utilized for each layer. It is essential to define specific fundamental parameters, including the optimizer (adam), loss function (MSE), and metrics (MSE). The model is trained using the training data. Optimal settings must be determined to improve the model's efficacy utilising a batch size of 128 and 40 epochs.

4.3.2. *Recurrent Neural Network (RNN) Model.* The RNN model comprises a configuration of five layers. The RNN has four hidden layers and a single output layer [29]. Table (7) presents the RNN hidden layers together with the corresponding units for each layer. To get knowledge, it is essential to define foundational parameters for the development of this model. The fundamental parameters employed are optimiser = "adam", loss = "MSE", and metrics = "MSE". The model uses a batch size of 128 and a total of 40 epochs.

4.3.3. *Long Short-Term Memory (LSTM).* Four layers make up the LSTM model. Three concealed and one output layers make up the four layers [30]. The next LSTM layer receives the output of the first. Second layer output becomes third layer input, while third layer output becomes fourth layer input. The final LSTM layer forecasts oil and gas output. Many trials determined the optimal number of hidden layers and neurones for the LSTM model. Verifying accuracy after each change showed that the four hidden layers worked well. Table (8) lists the LSTM model's hidden layer nodes and activation

TABLE 7. The constituents of the hidden and output layers for the RNN model.

Number of hidden layers	Units of nodes	Activation function
First hidden layer	32	Tanh
Second hidden layer	16	Tanh
Third hidden layer	8	Tanh
Forth hidden layer	4	Tanh
Fifth layer (output)	1	ReLU

TABLE 8. The constituents of the hidden and output layers for the LSTM model.

Number of hidden layers	Units of nodes	Activation function
First hidden layer	46	ReLU
Second hidden layer	46	ReLU
Third hidden layer	32	ReLU
Forth hidden layer	1	ReLU

functions. Train the model with training data. Using batch size 128 and epoch 20, find model parameters that improve performance.

4.3.4. *1-D CNN-based Regressor.* Convolutional Neural Networks (CNNs) are recognised for their resilience and have established themselves as the standard in several computer vision applications [31]. A distinctive characteristic that enhances the efficiency of CNNs in supervised learning is the spatial-local connection, which enables layers to communicate parameters [32]. Feature extraction in CNNs is predominantly dependent on convolution (Conv) layers, which execute convolution operations on the input data or feature map(s) utilising predefined kernels. The hyperparameters, including the quantity of hidden layers, kernel size (K), number of filters (F), subsampling factor, and activation function type employed in each layer, dictate the architecture of the 1-D CNN model. This convolution procedure will produce a volume of learnt feature maps. Given that the input consists of one-dimensional sequential data, the input convolutional layer processes a one-dimensional input sequence, $x(n) \in \mathbb{R}^{1\times 6}$. A convolution between the kernel, w(n), and the input produces a feature map, z(n) [33].

5. Experimental Result.

5.1. **Hardware and Software Requirements.** Processor: 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.42 GHz, System Type: 64-bit operating system, x64-based processor. RAM: 12.0 GB (11.8 GB usable).Operating System: Windows 10 Home. Our study used the Python programming languages Anaconda Jupyter and Collaboration Online (https://colab.research.google.com) as our Python environments.

- NumPy: A package that enables array manipulation in Python.
- Tensor flow and Keras: These libraries make writing code for Deep Learning models easier.
- Scikit-learn (Sklearn): used for partitioning and fitting datasets into numerous machine learning models.
- Pandas: This library is employed for importing and partitioning datasets.

5.2. **Evaluation Metrics.** The evaluation metrics used to measure how well a model predicts continuous values, we utilized:

• Mean Squared Error (MSE) for large errors especially bad and should be penalized more.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

Where: n: Number of data points, y_i: Actual value at index i, $\hat{y}_i$: Predicted value at index i, $(y_i - \hat{y}_i)^2$: Squared error for each prediction,

• Mean Absolute Error (MAE) to easyly interpret an error metric with less affection from the outliers.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

Where: n : Number of data points, y_i : Actual value at index i, $\hat{y}_i$: Predicted value at index i, $|y_i - \hat{y}_i|$: Absolute error for each prediction.

• R-squared / Coefficient of Determination (R^2) to understand the models performance for dataset variety.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

Where: y_i : Actual value at index i, $\hat{y}_i$: Predicted value at index i, $\bar{y}$: Mean of all actual values, $(y_i - \hat{y}_i)^2$: Squared error for each prediction (model error), $(y_i - \bar{y})^2$: Total variance in the data.

5.3. **The Result and Discussion.** The result of ML models with different parameters in oil and gas production forecasting output, as shown in Table (9). The mentioned parameters are essential in influencing ML models' performance. For testing all ML supervised regression models (DTR, RFR, KNN, and XGBoost).

Table 9. Machine learning models parameters.

Model	Parameter	Value
DTR	Max_depth	500
	Random_state	33
RFR	n_estimators	50
	Max_depth	15
	Random_state	33
KNN	n_neighbors	13
XGBoost	n_estimators	30
	seed	300

The result of ML model's performance using the evaluation matrix MSE, R^2, and MAE, see table (10).

According to the results, the KNN model performs best with low MAE and MSE values and R^2 values close to one, indicating a good correlation between actual and predicted values. The ML models (RFR, DTR, KNN, and XGBoost) closely match anticipated and actual oil output Figure (4). Effective oil production forecasting models have lower MAE, MSE, and higher R^2 values that match projected and actual values.

The result of DL models performance using the evaluation metrics (MSE, R^2, and MAE) presented in table (12), performance values for the employed DL models for predicting oil and gas. In all model the values were superior and quite promising, a comparison between the DL and ML models demonstrated in table (13) based on the value of R^2, a graphical

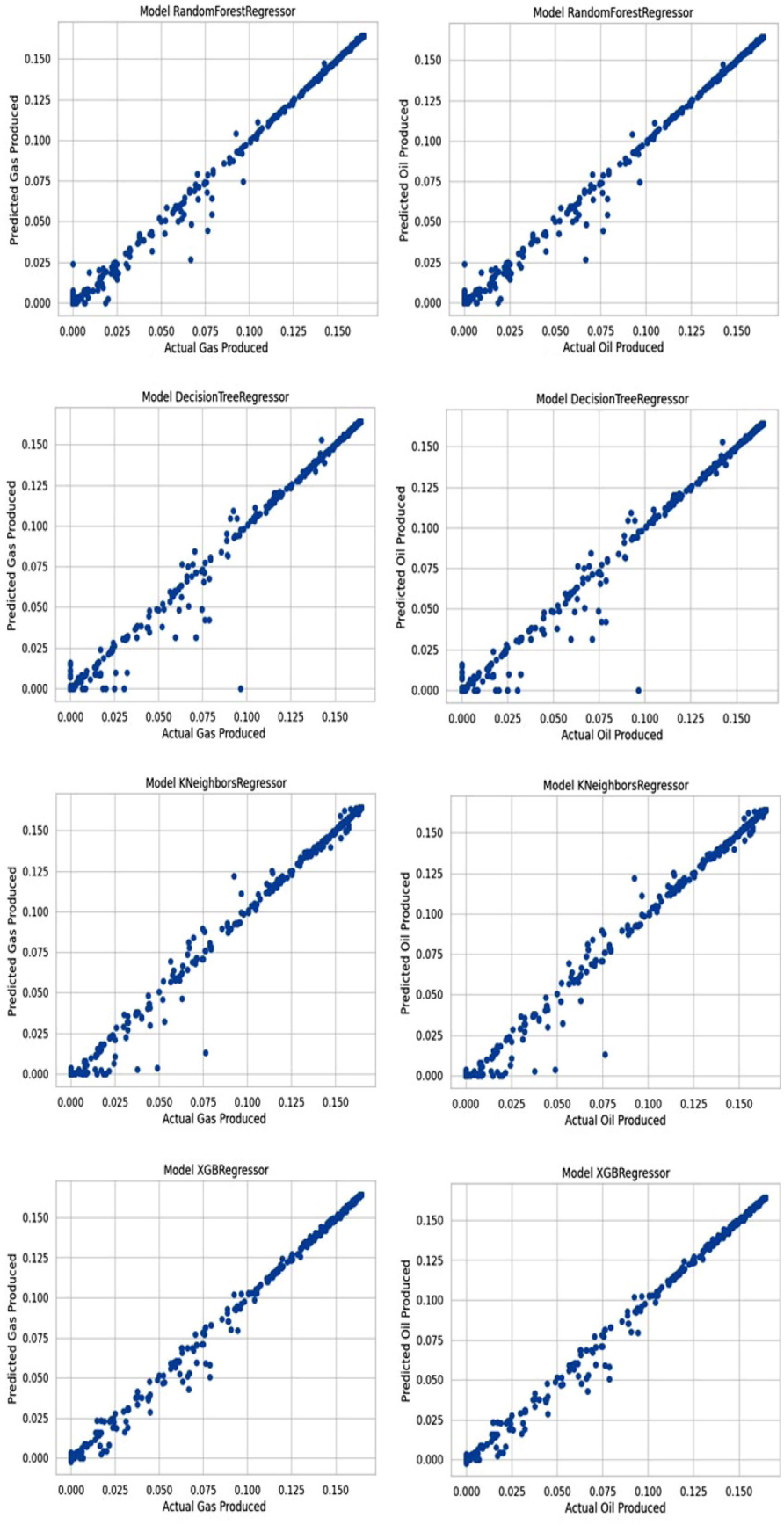

FIGURE 4. Machine Learning Predicted oil and gas produce VS. Actual produces.

TABLE 10. Evaluation Metrics for ML models Performance.

Model	Output	MAE	MSE	R^2
RFR	Oil	0.0014	0.0002	0.9936
	Gas	0.0017	0.0002	0.9982
DTR	Oil	0.0017	0.0003	0.9893
	Gas	0.0016	0.0003	0.9972
KNN	Oil	0.0015	0.0002	0.9937
	Gas	0.0012	0.0001	0.9988
XGBoost	Oil	0.0020	0.0002	0.9921
	Gas	0.0027	0.0004	0.9969

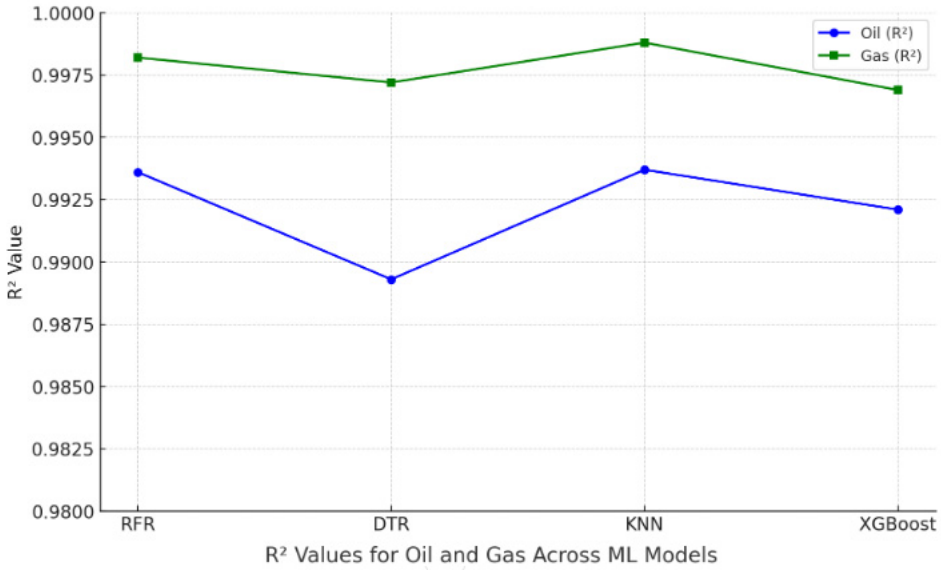

FIGURE 5. The R^2 curve for Oil and Gas across ML models.

chart in figure (7) show the comparison in R^2 value across all the models. For extend the performance generalization, several of most recent studies compared with the proposed models refereeing to the dataset used and the values of R^2 listed in table (14).

5.4. **limitation.** Although deep learning models like ANN, RNN, LSTM, and 1-D CNN achieved high predictive accuracy, they present interpretability challenges, which can limit user trust and hinder decision-making in operational settings. In contrast, models such as DTR and KNN offer clear reasoning paths, making them more suitable when explainability is essential. Additionally, deploying these models in real-world oil and gas fields involves challenges such as sensor data inconsistency, missing values, changing production dynamics, and the necessity for periodic retraining to adapt to new operational conditions. While this dataset is well-documented and valuable for benchmarking, it may not reflect modern advancements in extraction technologies, sensor systems, or reservoir management practices. Consequently, future research should involve recent or real-time production data to enhance model applicability in contemporary oil and gas fields.

6. **Conclusion and Future Work.** ML and DL models for petroleum oil and gas production were described in this research. The study emphasizes forecasting oil and gas production for informed decision-making, reserve estimation, production optimisation, market dynamics, trend prediction, project commercial viability analysis, and recovery

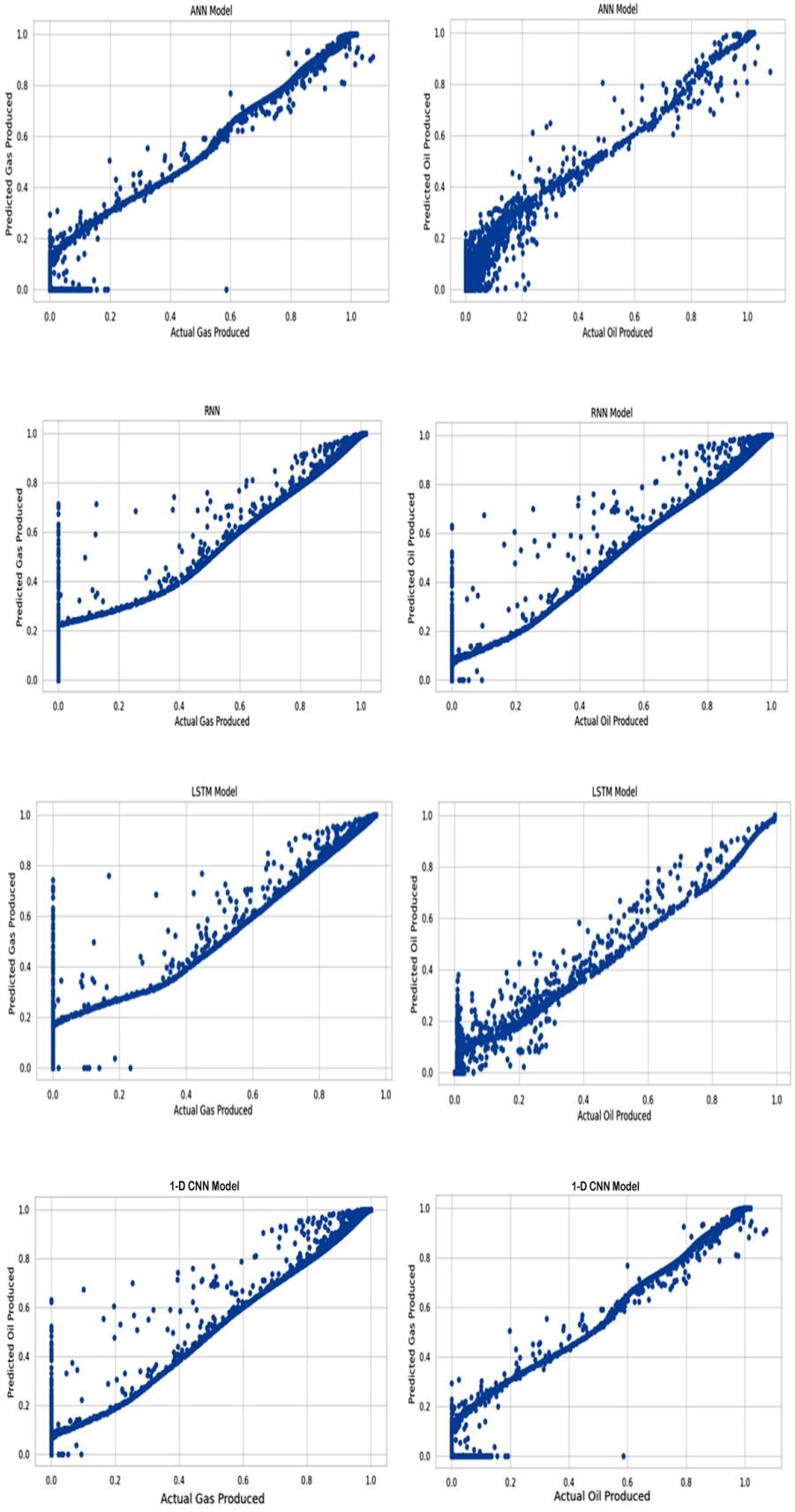

FIGURE 6. Deep Learning Predicted oil and gas produce VS. actual produce.

TABLE 11. The constituents of the hidden and output layers for the DL models.

Model	Layer Type	Number of Nodes (Units)	Activation Function
ANN	First Hidden Layer	64	Tanh
	Second Hidden Layer	128	Tanh
	Third Layer (Output)	1	ReLU
RNN	First Hidden Layer	32	Tanh
	Second Hidden Layer	16	Tanh
	Third Hidden Layer	8	Tanh
	Fourth Hidden Layer	4	Tanh
	Output Layer	1	ReLU
LSTM	First Hidden Layer	64	ReLU
	Second Hidden Layer	64	ReLU
	Third Hidden Layer	32	ReLU
	Output Layer	1	ReLU
1-D CNN	Input Layer	(32, 5, 12)	None
	Conv1D (L1)	64	ReLU
	Conv1D (L2)	64	ReLU
	Conv1D (L3)	64	ReLU
	MaxPooling1D (L4)	64	ReLU
	Flatten (L5)	64	None
	Dense (L6)	1	Linear

TABLE 12. Evaluation Metrics for DL models Performance.

Model	Output	MAE	MSE	R^2
ANN	Oil	0.0020	0.0004	0.9545
	Gas	0.0060	0.0005	0.9951
RNN	Oil	0.0070	0.0011	0.9894
	Gas	0.0040	0.0005	0.9949
LSTM	Oil	0.1417	0.0003	0.9667
	Gas	0.0084	0.0011	0.9893
1-D CNN	Oil	0.0060	0.0005	0.9951
	Gas	0.0070	0.0011	0.9894

TABLE 13. The result of ML VS. DL models.

Model	Oil R^2 Value	Gas R^2 Value
RFR	0.9936	0.9982
DTR	0.9892	0.9972
KNN	0.9937	0.9988
XGBoost	0.9921	0.9969
ANN	0.9545	0.9951
RNN	0.9894	0.9949
LSTM	0.9667	0.9893
1-D CNN	0.9951	0.9894

rate, cost, and operational efficiency. We use machine learning and deep learning to construct a more accurate and efficient production forecasting model than DCA and NRS. The main idea was to construct a simple and practical ML and DL model for speedy, informed decision-making. Eight production forecasting approaches employing ML and DL

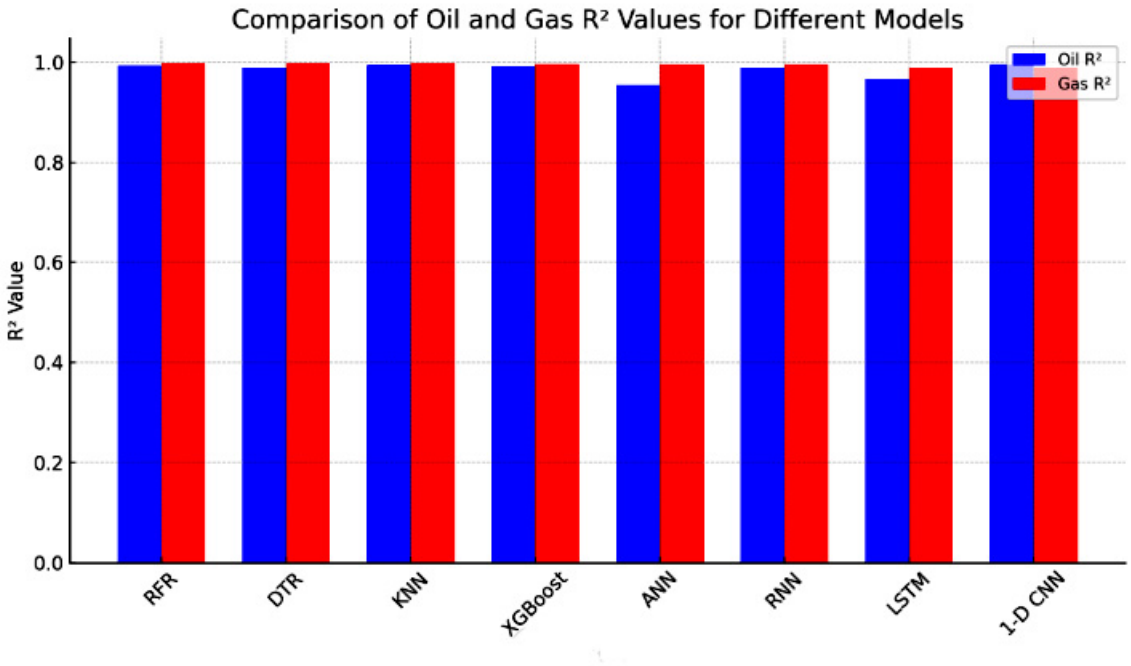

FIGURE 7. Comparing results between ML and DL models.

models were constructed and tested: DTR, KNN, RFR, XGBoost, ANN, RNN, LSTM, and 1-D CNN. To evaluate the models performance us apply the most popular metrics(MSE, MAE and R^2), these metrics calculate the prediction values for each modles, covering several cases such as dealing with large errors allwonig easy interret and less affected by outliers and to understant how well the models explains the variance in dataset. Models accurately anticipate oil and gas output. Data and model predictions are strongly correlated in the production. Understanding that the best model depends on the data is crucial. Therefore, testing different models and assessing their performance is essential to choose the best one. While this dataset is well-documented and valuable for benchmarking, it may not reflect modern advancements in extraction technologies, sensor systems, or reservoir management practices. Consequently, future research should involve recent or real-time production data to enhance model applicability in contemporary oil and gas fields. Future works will aim to employ more specialized datasets, concentrating on specific fields or wells and the variables affecting their output. Employing supplementary models like VGG-19 with the previously utilized approaches.

TABLE 14. Compares proposed models result with previous studies.

Study	Dataset	Technique	R^2 (Oil)	R^2 (Gas)
C. Tan et al. (2021) [15]	137 fractured wells, Sichuan, China	RF	0.72	N/V
		BP	0.82	N/V
		SVR	0.74	N/V
		XGBoost	0.87	N/V
		Light GBM	0.83	N/V
		LR	0.78	N/V
N. M. Ibrahim et al. (2022) [17]	Saudi Aramco	MLR	0.834	0.7684
		PLR	0.966	0.9185
		SVR	0.9659	0.9185
		DTR	0.9225	0.9236
		RFR	0.9355	0.9247
		XGBoost	0.9561	0.9336
		ANN	0.9697	0.9185
		RNN	0.9785	0.8787
L. Mai-Cao et al. (2022) [18]	Oilfield X, Southern Vietnam	RF	0.87	N/V
		GB	0.85	N/V
		KNN	0.91	N/V
		SVR	0.96	N/V
		MLP	0.93	N/V
		CNN	0.88	N/V
		LSTM	0.90	N/V
		GRU	0.92	N/V
A. E. Al-Aghbari et al. (2022) [19]	Volve oil field database	GA-TCN-LSTM	0.93	N/V
		TCN	0.92	N/V
		LSTM	0.91	N/V
		GRU	0.92	N/V
		RNN	0.92	N/V
W. Liu, Z. Chen et al. (2023) [20]	Actual reservoirs	ANN	0.795	N/V
		XGB	0.857	N/V
S. Hosseini et al. (2022) [22]	Volve oil field dataset	LSTM	0.97–0.98	N/V
		CNN	0.84–0.94	N/V
Proposed Models (2025)	USA wells, New York State	RFR	0.9936	0.9982
		DTR	0.9892	0.9972
		KNN	0.9937	0.9988
		XGBoost	0.9921	0.9969
		ANN	0.9545	0.9951
		RNN	0.9894	0.9949
		LSTM	0.9667	0.9893
		1-D CNN	0.9951	0.9894

REFERENCES

[1] Y. Cheraghi, S. Kord, V. Mashayekhizadeh, Application of machine learning techniques for selecting the most suitable enhanced oil recovery method; challenges and opportunities, *Journal of Petroleum Science and Engineering*, vol. 205, 108761, 2021.

[2] Z. N. Nemer, Oil and Gas Production Forecasting Using Decision Trees, Random Forest, and XG-Boost, *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 16, no. 1, pp. 9–20, 2024.

[3] T. Doan, M. Van Vo, Using machine learning techniques for enhancing production forecast in north Malay Basin, in *International Field Exploration and Development Conference*, Singapore, Springer Singapore, pp. 114–121, 2020.

[4] A. M. AlRassas, M. A. Al-Qaness, A. A. Ewees, S. Ren, M. Abd Elaziz, R. Damaševičius, T. Krilavičius, Optimized ANFIS model using Aquila Optimizer for oil production forecasting, *Processes*, vol. 9, no. 7, 1194, 2021.

[5] I. Makhotin, D. Orlov, D. Koroteev, Machine learning to rate and predict the efficiency of waterflooding for oil production, *Energies*, vol. 15, no. 3, 1199, 2022.

[6] M. A. Al-Qaness, A. A. Ewees, L. Abualigah, A. M. AlRassas, H. V. Thanh, M. Abd Elaziz, Evaluating the applications of dendritic neuron model with metaheuristic optimization algorithms for crude-oil-production forecasting, *Entropy*, vol. 24, no. 11, 1674, 2022.

[7] R. de Oliveira Werneck et al., Data-driven deep-learning forecasting for oil production and pressure, *Journal of Petroleum Science and Engineering*, vol. 210, 109937, 2022.

[8] E. H. Alkhammash, An optimized gradient boosting model by genetic algorithm for forecasting crude oil production, *Energies*, vol. 15, no. 17, 6416, 2022.

[9] C. S. Ng, A. J. Ghahfarokhi, M. N. Amar, Well production forecast in Volve field: Application of rigorous machine learning techniques and metaheuristic algorithm, *Journal of Petroleum Science and Engineering*, vol. 208, 109468, 2022.

[10] W. Li, Z. Dong, J. W. Lee, X. Ma, S. Qian, Development of decline curve analysis parameters for tight oil wells using a machine learning algorithm, *Geofluids*, vol. 2022, 8441075, 2022.

[11] Z. Tariq et al., A systematic review of data science and machine learning applications to the oil and gas industry, *Journal of Petroleum Exploration and Production Technology*, 2021.

[12] Davies, Jasmy and Sivakumari, S, A Comparative Analysis of Destructive Methods and Non-Destructive Methods with Machine Learning and Deep Learning Approaches for Rice Leaf Disease Identification, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 15, No. 2, pp. 87–97, 2024.

[13] B. M. Negash, A. D. Yaw, Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection, *Petroleum Exploration and Development*, vol. 47, no. 2, pp. 383–392, 2020.

[14] S. Choubey, G. P. Karmakar, Artificial intelligence techniques and their application in oil and gas industry, *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3665–3683, 2021.

[15] C. Tan, J. Yang, M. Cui, H. Wu, C. Wang, H. Deng, W. Song, Fracturing productivity prediction model and optimization of the operation parameters of shale gas well based on machine learning, *Lithosphere*, 2021, Special 4, 2884679.

[16] G. Hui, S. Chen, Y. He, H. Wang, F. Gu, Machine learning-based production forecast for shale gas in unconventional reservoirs via integration of geological and operational factors, *Journal of Natural Gas Science and Engineering*, vol. 94, 104045, 2021.

[17] N. M. Ibrahim et al., Well performance classification and prediction: deep learning and machine learning long term regression experiments on oil, gas, and water production, *Sensors*, vol. 22, no. 14, 5326, 2022.

[18] L. Mai-Cao, H. Truong-Khac, A comparative study on different machine learning algorithms for petroleum production forecasting, *Improved Oil and Gas Recovery*, 2022.

[19] A. E. Al-Aghbari, B. K. Lee, Multivariate Time Series Forecasting of Oil Production Based on Ensemble Deep Learning and Genetic Algorithm, Available at SSRN 4460174, 2022.

[20] W. Liu, Z. Chen, Y. Hu, L. Xu, A systematic machine learning method for reservoir identification and production prediction, *Petroleum Science*, vol. 20, no. 1, pp. 295–308, 2023.

[21] X. Y. Wang, Y. J. Ma, E. Z. Fei, Y. F. Gao, Daily production prediction of oil wells based on machine learning, in *International Conference on Automation Control, Algorithm, and Intelligent Bionics (ACAIB 2023)*, SPIE, vol. 12759, pp. 516–520, 2023.

[22] S. Hosseini, T. Akilan, Advanced deep regression models for forecasting time series oil production, arXiv preprint arXiv:2308.16105, 2023.

[23] Z. N. Nemer, Oil and Gas Production Forecasting Using Decision Trees, Random Forest, and XGBoost, *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 16, no. 1, pp. 9–20, 2024.

[24] J. Wang, K. Sain, X. Wang, N. Satyavani, S. Wu, Characteristics of bottom-simulating reflectors for Hydrate-filled fractured sediments in Krishna–Godavari basin, eastern Indian margin, *Journal of Petroleum Science and Engineering*, vol. 122, pp. 515–523, 2014.

[25] A. Liaw, M. Wiener, Classification and regression by randomForest, *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[26] S. Zhou, X. Xie, Application of K-nearest neighbor algorithm for prediction of oil and gas reservoir properties, *Journal of Petroleum Science and Engineering*, vol. 177, pp. 845–852, 2019.

[27] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[28] G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[29] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555, 2014.

[30] S. A. Fattah, M. R. H. A. M. Al-Harthy, S. B. M. B. Al-Saidi, Application of Long Short-Term Memory (LSTM) for forecasting oil and gas production in reservoir management, *Journal of Petroleum Science and Engineering*, vol. 187, 106684, 2020.

[31] T. Akilan, Q. J. Wu, A. Safaei, J. Huo, Y. Yang, A 3D CNN-LSTM-based image-to-image foreground segmentation, *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 959–971, 2019.

[32] S. Hosseini, A. Shahbandegan, T. Akilan, Deep neural network modeling for accurate electric motor temperature prediction, in *2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 170–175, 2022.

[33] J. Zhao, X. Mao, L. Chen, Speech emotion recognition using deep 1D & 2D CNN LSTM networks, *Biomedical Signal Processing and Control*, vol. 47, pp. 312–323, 2019.