Journal of Farayandno – Vol. 20 – No. 90 (2025): pp. 73-90

National Iranian Oil Refining and Distribution Company (NIORDC)

Journal of Farayandno

Research Paper

DOI: 10.22034/farayandno.2025.2069800.2012

Application of Artificial Intelligence Algorithms for Predicting and Optimizing Rate of Penetration in Shadegan Oil Field Formations Using Hydraulic and Formation Pressure Data

Mahdi nazarisaram *[1], **Rahim Dabiri** [2]

[1] Department of petrolume mining and Materials, CT.C, Islamic Azad University, Tehran, Iran
[2] Department of Geology, Ma.C., Islamic Azad University, Tehran, Iran

Received: 31 May 2025 **Accepted:** 3 Aug 2025

1. ABSTRACT

Accurate prediction of the Rate of Penetration (ROP) remains a major challenge in drilling optimization due to its highly nonlinear and multivariate nature. In this study, real field data from the Shadegan oil field including hydraulic parameters (e.g., pump pressure, flow rate), formation pressure, and operational variables were used to develop and compare two deep learning models: Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). After preprocessing and time-series segmentation, models were evaluated using RMSE, MAE, and R^2 metrics. The LSTM model demonstrated superior performance with $R^2 = 0.95$, MAE = 0.0236, and RMSE = 0.0346, significantly outperforming the RNN ($R^2 = 0.85$). Furthermore, the trained LSTM model was employed as a surrogate for ROP optimization, yielding an optimal ROP of 8.63 m/h more than 3.3 times the observed field average (2.6 m/h) under safe operational constraints. This data-driven approach offers a practical framework for real-time drilling advisory systems.

Keywords: Rate of Penetration (ROP), Shadegan oil Field, Artificial Intelligence (AI), LSTM Network, Hydraulic Data, Drilling Optimization, Deep Learning

2. INTRODUCTION

ROP is a key indicator of drilling efficiency, directly affecting time and cost. However, it is influenced by complex, nonlinear interactions among geological (e.g., rock strength), hydraulic (e.g., mud flow), and operational factors (e.g., WOB, RPM). Traditional empirical models often fail to capture these dynamics. Deep learning, especially LSTM networks, has shown promise in modeling time-dependent, nonlinear drilling data [1]. This study applies LSTM to real Shadegan field data including formation pressure and hydraulic variablesfor the first time, offering a novel and practical contribution to drilling optimization in complex carbonate and evaporite formations.

While recent studies have employed ANN, SVM, or even LSTM for ROP prediction [1–4], none have incorporated real-time formation pressure as a direct input despite its critical role in controlling borehole stability and bit-rock interaction, especially in overpressured or reactive formations like Gachsaran. Moreover, most prior works rely on simulated or synthetic data, or omit hydraulic-transient effects that dominate in deep, high-pressure wells. This study bridges this gap by leveraging high-frequency field measurements from vertical wells in Shadegan, where pore pressure gradients and lithological transitions between carbonate and evaporite layers introduce significant ROP variability.

* mahdinazarisaram@iauctb.ac.ir

Please Cite This Article Using:
Nazari Saram, M., Dabiri, R., "Application of Artificial Intelligence Algorithms in Predicting and Optimizing Rate of Penetration in the Shadegan Oilfield Formations Based on Hydraulic and Formation Pressure Data", Journal of Farayandno – Vol. 20 – No. 90, pp. 73-90, In Persian, (2025).

3. MATERIALS AND METHODS

High-frequency drilling data were acquired from three vertical wells drilled in the Shadegan oil field, southwestern Iran, spanning depths from 1,200 to 3,200 meters true vertical depth (TVD). The dataset encompasses over 30,000 time-ordered records collected during drilling of two geologically distinct formations: the Asmari Formation (a heterogeneous carbonate reservoir with variable porosity and UCS) and the Gachsaran Formation (an evaporitic sequence dominated by interbedded gypsum, anhydrite, and halite layers, known for its pressure sensitivity and borehole instability risks).

The input feature vector comprised 12 operational and geomechanical parameters, including:

- Mechanical drilling parameters: Weight on Bit (WOB, klbf), Rotary Speed (RPM), Surface Torque (lb·ft);
- Hydraulic variables: Pump Pressure (psi), Flow Rate (GPM);
- Geomechanical properties: Unconfined Compressive Strength (UCS, MPa), Young's Modulus (GPa), Poisson's Ratio;
- In-situ formation pressure (psi), derived from real-time pore pressure monitoring a novel inclusion in ROP modeling for this field.

The target variable was the Rate of Penetration (ROP, m/h), measured continuously via surface logging-while-drilling (LWD) sensors.

- Prior to modeling, the dataset underwent rigorous preprocessing:
- Outlier removal using the Interquartile Range (IQR) method and $\pm3\sigma$ statistical thresholds;
- Gap-filling for missing values via a 5-point moving average to preserve temporal continuity;
- Feature scaling using Min-Max normalization (range: [0,1]) to maintain physical interpretability of bounded operational parameters (e.g., WOB $\geq$ 0, flow rate $\geq$ 0);
- Temporal sequence generation via a sliding window approach, with window lengths of 12 timesteps for LSTM and 10 for RNN, selected based on autocorrelation decay analysis of ROP residuals to capture relevant memory effects in drilling dynamics.

Two deep recurrent architectures were implemented in TensorFlow 2.12/Keras:

- A standard RNN with 50 tanh-activated recurrent units;
- A stacked LSTM with 50 memory cells, sigmoid gates (input/forget/output), and ReLU activation in the output layer.

Both models shared identical hyperparameters:

- Loss function: Mean Squared Error (MSE);
- Optimizer: Adam (learning rate = 0.0008, β_1 = 0.9, β_2 = 0.999);
- Batch size: 64;
- Training epochs: up to 100, with early stopping (patience = 10) triggered by validation loss stagnation to mitigate overfitting.

Critically, the dataset was partitioned chronologically not randomly into 70% training, 15% validation, and 15% testing subsets to prevent temporal data leakage, a common pitfall in time-series modeling that artificially inflates performance metrics. This split ensures that the model is evaluated on future, unseen drilling conditions, enhancing its operational realism and deployment readiness in real-time advisory systems. To prevent temporal data leakage, the dataset was split chronologically (not randomly): the first 70% of time-ordered samples for training, next 15% for validation, and final 15% for testing. Min-Max scaling was preferred over Z-score normalization to preserve physical bounds of operational parameters (e.g., WOB $\geq$ 0, flow rate $\geq$ 0). The sliding window size (12 for LSTM) was selected based on autocorrelation analysis of ROP residuals, ensuring capture of relevant memory effects in drilling dynamics. Model training was implemented in TensorFlow 2.12/Keras, with early stopping triggered if validation loss did not improve for 10 consecutive epochs.

4. RESULTS AND DISCUSSION

Correlation analysis revealed strong positive relationships between ROP and key operational parameters specifically WOB (r = 0.58) and flow rate (r = 0.51) highlighting their dominant influence on drilling efficiency. The LSTM model significantly outperformed the RNN, achieving an R^2 of 0.95 versus 0.85, along with markedly lower prediction errors (MAE = 0.0236, RMSE = 0.0346). SHAP (SHapley Additive exPlanations) analysis further identified formation pressure as the second-most influential feature contributing ~12% to ROP variance particularly within Gachsaran evaporite intervals, where pore pressure spikes induced ROP reductions of up to 40%. The LSTM accurately captured sharp ROP transients across lithological boundaries, whereas the RNN exhibited larger residuals and outliers due to its limited capacity to model long-term temporal dependencies. Residual diagnostics confirmed that over 80% of LSTM prediction errors fell within ±0.1 m/h, with a near-Gaussian distribution ($\mu \approx 0$, $\sigma = 0.028$), indicating robust and unbiased performance. In contrast, the RNN displayed heteroscedastic error behavior in evaporite zones, reflecting its inadequacy in handling lithology-dependent drilling dynamics. When benchmarked against the empirical Warren model (R^2 = 0.62), the LSTM reduced MAE by 71%, underscoring its superior ability to represent complex, nonlinear bit–rock–hydraulic interactions. Leveraging the trained LSTM as a differentiable surrogate model, operational optimization yielded a predicted ROP of 8.63 m/h more than 3.3 times the field average (2.6 m/h) at WOB = 21.5 klbf, RPM = 138, and flow rate = 111 GPM. Critically, all optimized parameters remained within established safe operating envelopes, avoiding risks such as stick-slip, borehole instability, or excessive bit wear. These results validate the proposed framework as a high-fidelity, field-deployable solution for real-time ROP prediction and drilling optimization in complex carbonate-evaporite sequences.

5. CONCLUSION

The LSTM-based framework successfully predicted and optimized the Rate of Penetration (ROP) using high-fidelity, real-time field data from the Shadegan oil field marking the first integration of in-situ formation pressure alongside hydraulic variables for ROP modeling in this field. The model captured over 95% of the observed ROP variance ($R^2 = 0.95$), demonstrating exceptional fidelity in complex carbonate-evaporite sequences. Through constrained optimization, it identified operational parameters that yielded an ROP of 8.63 m/h more than 3.3 times the historical field average (2.6 m/h) while remaining within safe drilling envelopes. The architecture is readily deployable in real-time drilling advisory systems for dynamic decision support. Key limitations include reliance on field-specific data and the exclusion of downhole temperature effects; future work will investigate physics-informed neural networks and transfer learning for broader applicability.

6. REFERENCES

[1] Pei, Z.-J., et al., Machine learning applications in reservoir characterization, Pet. Sci., 21, pp. 582–596, 2024.
[2] Brenjkar, E., et al., Optimization of hydraulic fracturing parameters in unconventional reservoirs, J. Pet. Explor. Prod. Technol., Vol. 11, pp. 805–818, 2021.
[3] Al-AbdulJabbar, A., et al., Sustainable approaches in oilfield operations, Sustainability, 12, pp. 19, 2020.
[4] Abbas, A. K., et al., Thermodynamic modeling of multiphase flow in porous media, J. Energy Resour. Technol., 141, pp. 11, 2019.