

Research Article

Predicting Liquid Loading in Gas Condensate Wells Using Machine Learning to Enhance Production Efficiency

Ahmad Muhammad Salisu^{*} , Ibrahim Ayuba , Abdulrahman Abdulrasheed ,
Abubakar Usman 

Department of Petroleum Engineering, Faculty of Engineering and Engineering Technology, Abubakar Tafawa Balewa University, Bauchi, Nigeria

Abstract

Liquid loading in gas condensate wells drastically lowers gas production and increases operating expenses if unmanaged. The traditional empirical model often has difficulty representing the complex behaviours of multiphase flow and typically rely solely on historical data. In contrast, this study introduces a novel machine learning approach using a non-linear regression that integrates both historical and live data to predict liquid loading events in gas condensate wells with greater precision and adaptability. The newly developed machine learning Algorithm exhibited a very significant performance achieving an RMSE of 1.1293Mscf/d, MSE of 1.561 and R2 of 0.9978. The results surpass other machine learning approaches including the hybrid model with an RMSE of 2.8639 and R2 of 0.9978 and the Feed forward neural network, which have the value of R2 of 0.9833 respectively. The model's streamlined architecture requires moderate data volume and low computational power making it suitable for real time monitoring and seamless integration into digital oil field systems which improves usability. Also, its accuracy relies on high-quality data input, highlighting the importance of a strong sensor network. With lower computing power requirements and the ability to adjust to different field conditions, this makes it a practical, scalable tool and a cost-effective solution that improves decision-making in oil and gas field operations through insight based on data. This dual data driven approach offers a practical advancement over existing models, significantly contributing to the optimization of hydrocarbon recovery.

Keywords

Machine Learning, Gas-well, Liquid-loading, Liquid Prediction, Gas-condensate, Efficiency

1. Introduction

Natural gas remains a key part of the global energy mix because it has a relatively low carbon footprint and sees increasing demand in both industries and homes. One of the challenges in gas production is liquid loading. This problem is especially common in older gas and gas condensate wells. The liquid loading normally happens in the wellbore when fluids

like water and condensation build-up. This problem pointedly relegates the well productivity by slowing down the gas flow [1, 2]. This phenomenon not only hampers the rate s of production but also increases operational costs due to the need for frequent solutions like liquefaction and non-natural lifting systems [3].

^{*}Corresponding author: Ahmadsalisumuhammad3232@gmail.com (Ahmad Muhammad Salisu)

Received: 21 June 2025; Accepted: 4 July 2025; Published: 30 July 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Conventionally, the prediction and management of liquid loading highly relied on empirical correlations and mechanistic models such as Coleman correlation. Although these models have been essential, they fall short in representing the complex, nonlinear and dynamic behaviors of multiple flows in heterogeneous reservoir conditions [4]. However, these models involve many assumptions and simplifications, which limit their accuracy across varying field conditions.

Tree Based Models (XGBoost, Random Forest) are used for liquid loading classification and risk assessment due to their interpretability and handling of tabular data. For example, [5] applied XGBoost to predict loading onset using wellhead pressure, gas rate, and fluid gradients, achieving ~85% accuracy but requiring extensive feature engineering, its major drawback is Performance plateaus with high dimensional temporal data, and models often fail to generalize across wells with differing completion designs [6].

Deep Learning (LSTM, FFNN) are LSTMs have been employed to model temporal dependencies in sensor data (e.g., pressure transients, flow rates). [7] LSTMs demand large training datasets and are computationally intensive. FFNNs, while simpler, struggle with sequential data and require manual windowing of time series inputs [8].

Hybrid and Physics Informed ML integrates domain knowledge (e.g., conservation laws) with ML to improve generalizability. For instance, [9] combined a FFNN with mechanistic constraints, reducing prediction errors by 20% in low data regimes.

Gaps Addressed by This Study

Prior ML efforts face accuracy tradeoff, tree based models

excel on structured data but lack temporal sensitivity, while LSTMs require impractical data volumes for field deployment [5, 7].

Data Efficiency, most models are trained on historical datasets, limiting cross well applicability.

Few studies validate models in real-time operational settings. [8]'s LSTM, though accurate, had latency issues (>1 min per prediction), rendering it unsuitable for rapid decision making in real live.

The goals of this research work are to improve the gas well efficiency using machine learning technology to predict the liquid loading in gas condensate wells. To accomplish this, a new machine-learning algorithm was developed using Python Software. This will enhance accurate prediction through careful data processing, model training and optimization. After that, we will test and validate the model thoroughly with real production data to assess its reliability and generalizability. Finally, we will compare the algorithm's performance with current empirical and mechanistic models to demonstrate its potential to boost production rates, reduce operational costs, and extend well life.

2. Materials and Methods

2.1. Materials/Tools

Table 1 presents the materials/ tools used in this research work.

Table 1. Materials/Tools and their Purpose Used in this Research Work.

S/N	Equipment	Purpose
1.	PYTHON Software	For the development of new machine learning algorithms, modelling and simulation for accurately predicting liquid loading in gas condensate wells.
2.	Data collected from Gas condensate well.	Input and operating parameters for simulation (liquid flow rate and gas flow rate, etc.)
3.	Microsoft Excel	Data assembling and preparation

2.2. Methods

2.2.1. Data Processing and Analysis

In this study, the dataset came from gas condensate wells and includes important parameters such as gas flow rate (Qg), liquid flow rate (Ql), pressure drop across the liquid phase (ΔP_l), pressure drop across the gas phase (ΔP_g), minimum velocity (min), gas velocity (vg), liquid pressure (Pl), gas pressure (Pg), liquid density (ρ_l), gas density (ρ_g), and time (t).

Table 2 presents the operational and reservoir parameters important for modelling liquid accumulation in a gas condensate reservoir. However, these features are essential for modelling and training and verifying machine learning models and are very important for modelling and predicting the beginning of liquid loading. Also, these features serve as the basis for training and validating the machine learning models and are vital for accurately model predicting the commencement of liquid loading in a condensate gas well. The dataset was split into training mode carried 60%, the validation set held 10% and the testing phase held 30%. The testing

data set is used for model building; the validation set ensures that the model is built as optimally as possible while the testing is used to evaluate the model's final performance.

A statistical overview of 200 simulation samples used to predict the loading of liquid in a gas condensate well is shown in Table 2. Also, it highlights the distribution and variability of the input parameters feature and the target variable. However, the key variables such as ((gas flow rate (Q_g), liquid loading (Q_l) and liquid-to-gas density ratio (ρ_l/ρ_g)) demonstrate significant variability while showcasing a very

broad presentation of operational conditions. In comparison with the features of other operational parameters such as the condensate-gas ratio (CGR) and velocity ratio (v_{min}/v_g) exhibit limited variability, which may imply redundancy or need further transformation. The target variable representing the liquid loading shows an asymmetric distribution with a mean of 5.56 which supports regression-based modelling. The overall data set is well-balanced and diverse, providing a very strong foundation for future selection and applications of machine learning in reservoir analysis [7].

Table 2. Statistical Analysis for Collected Data.

Statistical term	Time	Qg (MMSCFD)	Ql (STB/D)	$\Delta P/\Delta P_g$	V min/vg	Pl/Pg	ρ_l/ρ_g	CGR	y
Simulation	200	200	200	200	200	200	200	200	200
mean	0.2827	5.0981	2.5714	2.8430	0.8873	0.8561	1.9960	0.1035	5.5623
std	0.2585	2.8015	1.4357	1.3829	0.1673	0.3424	0.6674	0.0553	1.2122
min	0.0029	0.5525	0.1248	0.5488	0.6109	0.3056	0.8140	0.0119	2.6464
25%	0.1001	2.6715	1.3812	1.6498	0.7416	0.3056	0.8140	0.0119	2.6464
50%	0.1794	5.1976	2.7540	2.8643	0.8803	0.8321	2.1081	0.1020	5.5633
75%	0.3954	7.6902	3.7368	4.1500	1.0346	1.1280	2.6032	0.1484	6.4350
max	0.9914	9.8754	4.9535	4.9987	1.1940	1.4962	2.9955	0.1995	8.4857

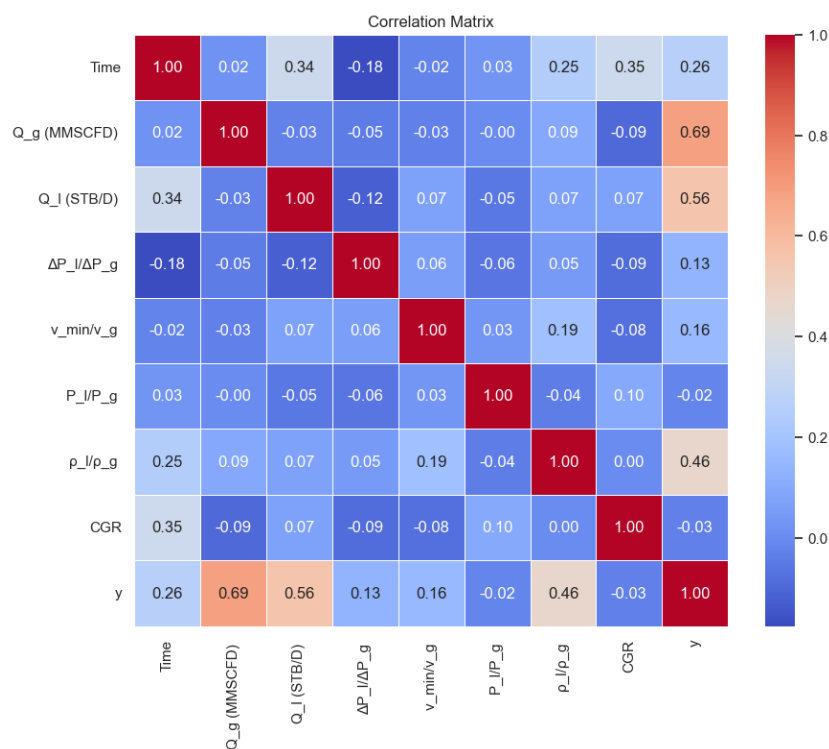


Figure 1. Pearson's Correlation is shown in the Heat Map of the Total Data set used for Liquid Loading Prediction in Gas Condensate Well.

2.2.2. Data Correlation Matrix Analysis

The Pearson correlation matrix heat map presented in Figure 2, highlights the linear relationship between the input and the target variable (y), which respects the liquid loading in the gas condensate wells. The variables such as gas flow rate (Q_g), liquid flow rate (Q_l), and gas density (ρ_g) demonstrate strong positive correlations with y , with a coefficient of 0.69.

0.56 and 0.46 respectively, this indicates a very good meaning in predicting liquid loading [8]. Also, in comparison with certain variables such as pressure drop ratio ($\Delta P_l/\Delta P_g$), velocity ratio (v_{min}/v_g), pressure ratio (P_l/P_g), CGR and time exhibit weak correlations with y , this may be due to limited linear predictive values when used independently. Though their low interdependency will help reduce multicollinearity [9].

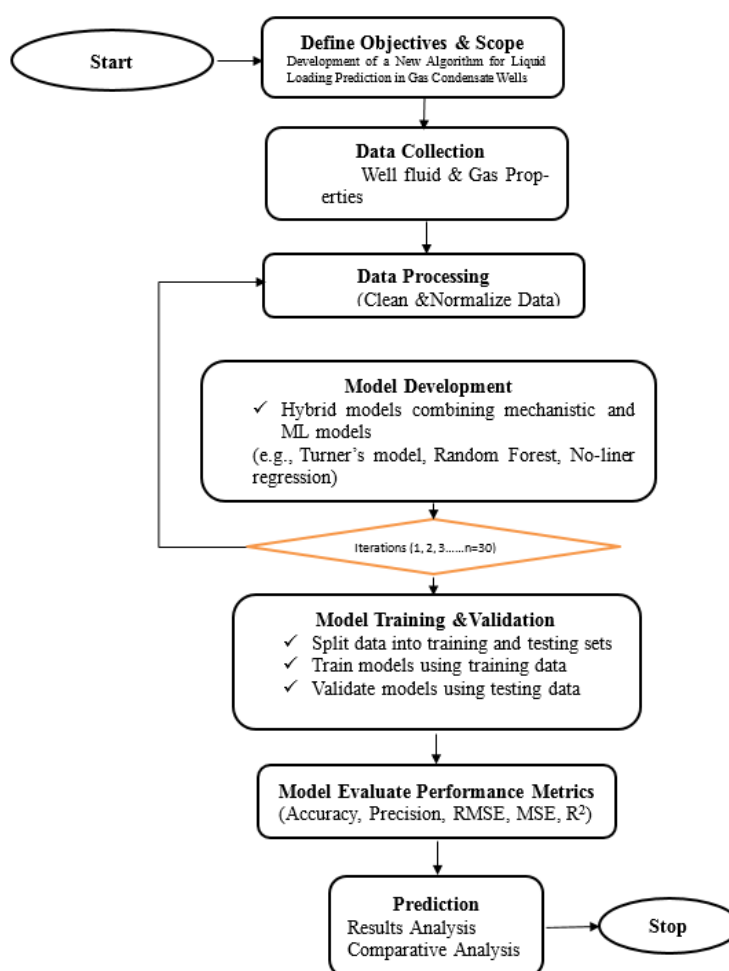


Figure 2. Conceptual Methodology Flowchart for developing a new Liquid Loading Prediction Algorithm in Gas Condensate Wells.

2.2.3. Model Development

The development of a new machine learning algorithm to predict liquid loading on gas condensate wells follows a systematic approach. The methodology integrates supervised machine learning (non-linear regression) with mechanistic simulation modelling enabling a robust comparison between the two methods. It is established by collecting details live data, including well pressure, temperature, production rates and the PVT properties. The data is then pre-processed to remove the outliers and generate key features such as superficial velocities and flow regime indicators. The developed machine learning algorithm uses a refined mechanistic model

such as a modification of Turner's critical velocity equation. The model accuracy is assessed using metrics like MSE, RMSE and R^2 based on historical shut-in and unloading events. The algorithm at the final stage is validated by comparing the simulated outputs to the actual well performances. Figures 4 and 5 depict the steps and results from the model development and validation.

2.2.4. Model Description

The machine learning (ML) method developed to predict liquid loading in the gas condensate is well presented in this section. In difference with the traditional Python-based simulation model, which is grounded in physical equations, the

ML model utilizes empirical data to identify patterns and predict the onset of liquid flooding. It learns the relationship between the input and out variables such as gas flow rate, pressure, velocity, and liquid loading event without relying explicitly on physical laws. The section also, covers the model architecture, feature engineering methods, algorithm selection, training and evaluation processes and includes a comparison with the conventional Python-based simulation approach.

2.2.5. Data Sources and Sampling Methods

The data for this study is sourced from operational records of gas condensate wells, obtained from industry partners and public oil and gas databases. These records provide real-time measurements collected during daily operations of the wells, covering both steady-state and transient conditions.

Data Collection Techniques

- 1) Field Data Field operators record real-time measurements of gas and liquid flow rates, pressures, and temperatures. These measurements are collected at regular intervals, ensuring that the dataset captures temporal variations in well conditions.
- 2) Production Logs: Production logs are used to track the performance of individual wells over time. These logs provide historical data on flow rates, pressure drops, and other operational parameters.
- 3) Reservoir Simulation Models: In cases where direct measurements are unavailable, reservoir simulation models may provide additional data, particularly for parameters like gas and liquid densities or wellbore velocities.
- 4) Sampling Frequency: The data is collected at different time intervals depending on the operational conditions of the well. For steady-state conditions, data might be sampled every few hours or days. During transient events or changes in production, more frequent sampling is employed to capture dynamic variations in the well performance.
- 5) Data Cleaning and Pre-processing: Before using the collected data in the PYTHON simulation or machine learning model, it must undergo a series of pre-processing steps to ensure its quality and suitability for analysis. These steps include cleaning the data, handling missing values, removing outliers, and normalizing or scaling the data.
- 6) Feature Scaling and Normalization: For the machine learning model, it is essential to scale and normalize the data, especially when the input features have different units and ranges. This step ensures that the model can learn efficiently and accurately.
- 7) Feature Engineering: Feature engineering is the process of creating new features or transforming existing ones to improve the performance of the machine learning model. For this study, new features might be derived from existing ones to capture interactions between variables. For example: *Ratio Features: Ratios such as (liquid-to-gas*

flow rate ratio) or (liquid-to-gas pressure ratio) can provide meaningful insights into the dynamics of liquid loading.

The basic assumptions used in this research work are as follows:

- 1) The equation presumes a gas-dominated flow, incorporating the liquid effect indirectly through the velocity term.
- 2) The well operates under transient flow conditions, where both pressure and flow rate vary over time rather than remaining constant.
- 3) The reservoir is assumed to have uniform permeability and porosity, reducing variability in flow behaviour.
- 4) It is assumed that there are no influences from nearby wells, fractures, or external pressure support that could alter the flow dynamic.
- 5) The rate of prediction decreases over time as indicated by following an exponential trend expressed as $e^{(-\alpha t)}$.
- 6) The constants (C , α , and β) are determined through curve fitting rather than strict physical derivation.
- 7) The production rate is modelled using pressure and velocity ratios, avoiding the complexity of full multiphase flow analysis.

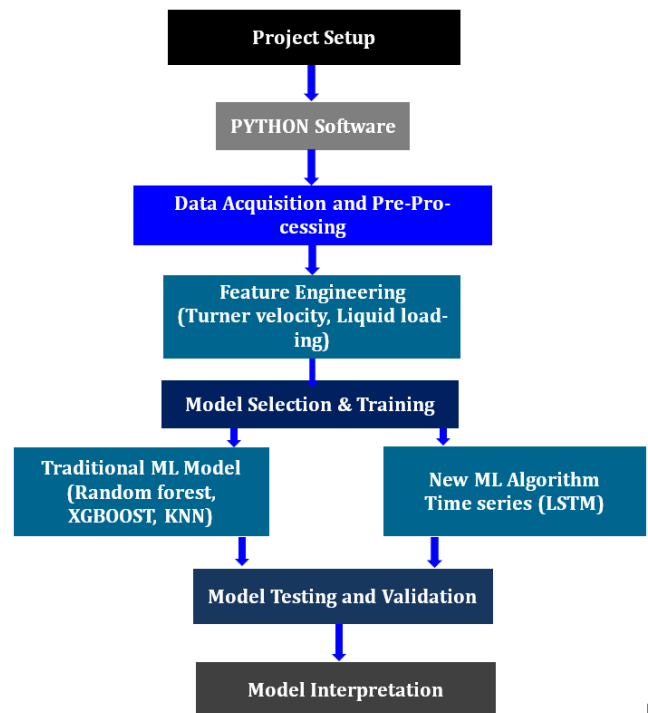


Figure 3. Python process flow chart for the development of a new algorithm for liquid loading prediction in gas condensate wells.

2.2.6. Mathematical Model Development

A mathematical frame that accurately represents the key physical processes in multiple-phase flow is required to develop a model for the prediction of liquid loading in petroleum reservoirs. This section presents the equations derivation of giving

flow in the well, incorporating transient flow assumptions and relevant boundary conditions. Implementing a transient approach, the model dynamically simulates the liquid loading over time, reflecting the evolving behaviours of the reservoir and wellbore system.

2.2.7. Transient Flow in Reservoir Engineering

The transient flow used in reservoir engineering refers to the time before the system stabilizes. The diffusion-like behavior controls fluid movement during this phase and pressure propagation through the reservoir is unstable. The diffusivity equation which derived from the equation for fluid flow in a porous medium is obtained by combining the equation of state (fluid properties), Darcy's Law (fluid flow in porous media) and the continuity equation (mass conservation) [10]:

The diffusivity Equation as expressed in Equation (1) for a single-phase flow in a homogenous reservoir [11]:

$$\frac{\partial P}{\partial t} = \frac{k}{\phi \mu c_t} \left(\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial P}{\partial r} \right) \right) \quad (1)$$

Where: $P = P(r,t)$ is pressure, a function of radial distance r and time t ; k = permeability; ϕ = porosity; μ = fluid viscosity; c_t = total compressibility.

Solving this partial differential equation (PDE) under certain boundary conditions leads to well-known solutions such as the exponential pressure decline in transient flow.

2.2.8. Exponential Pressure Decline in Transient Flow

For a well in an infinite-acting reservoir producing at a constant rate, the pressure solution can be expressed in Equation (2) [11]:

$$P(r, t) = P_i - \frac{qB\mu}{4\pi kh} E_i \left(-\frac{r^2}{4\alpha t} \right) \quad (2)$$

Where $E_i(x)$ is the exponential integral as shown in Equation (3).

$$\alpha = \frac{k}{\phi \mu c_t} \quad (3)$$

For small x , the exponential integral function is approximated by exponential decay as present in Equation (4)[4]:

$$E_i(-x) \approx e^{-x} \quad (4)$$

Thus, pressure decline follows an exponential decay as present in Equation (5):

$$P(r, t) \approx C \cdot e^{-at} \quad (5)$$

2.2.9. Modification to Include Pressure and Velocity Ratios

To improve the accuracy of transient flow predictions, the equation is modified to incorporate pressure and velocity ratios. The enhanced pressure and velocity ratios are given in Equation (6) [4]:

$$y = C \cdot P_{ratio} \cdot e^{-at} + \beta \cdot v_{ratio} \quad (6)$$

Where:

P_{ratio} = the pressure-related effect, modifying the transient pressure decline model; v_{ratio} = represents the velocity depending on the term that influences the production rate; β = is a coefficient that modulates the influences of the velocity term within the model. by implanting the curve fitting techniques such as least square regression, the model constant C , α and β are optimized using actual filed data. Thus, this enables the model to adapt dynamics to real-time production conditions, thereby improving the precision of liquid loading predictions.

2.2.10. Final Equation and Application in Simulation

The final transient flow in Equation (7) used in the simulation is:

$$y = C \cdot P_{ratio} \cdot e^{-at} + \beta \cdot v_{ratio} \quad (7)$$

The equation provides a practical approach for modeling time-dependent pressure decline without the need to solve complex partial differential equations (PDEs). It shows especially valuable performance in gas well analysis be enabled by engineers [12]:

- 1) To evaluate historical pressure behaviors to detect signs of liquid loading
- 2) Predict future well behaviors using real-time data inputs
- 3) Optimize the production parameters with tactical dynamic adjustment

Additionally, the facility's historical matching where actual data filed is used fine tune parameters (C , α , β) for generative predictive accuracy. By solving this equation, engineering gains critical insight into the transient pressure behavior of gas condensate wells. This predictive performance allows for early liquid loading detection and supports proactive decision-making to maintain optimum well performance.

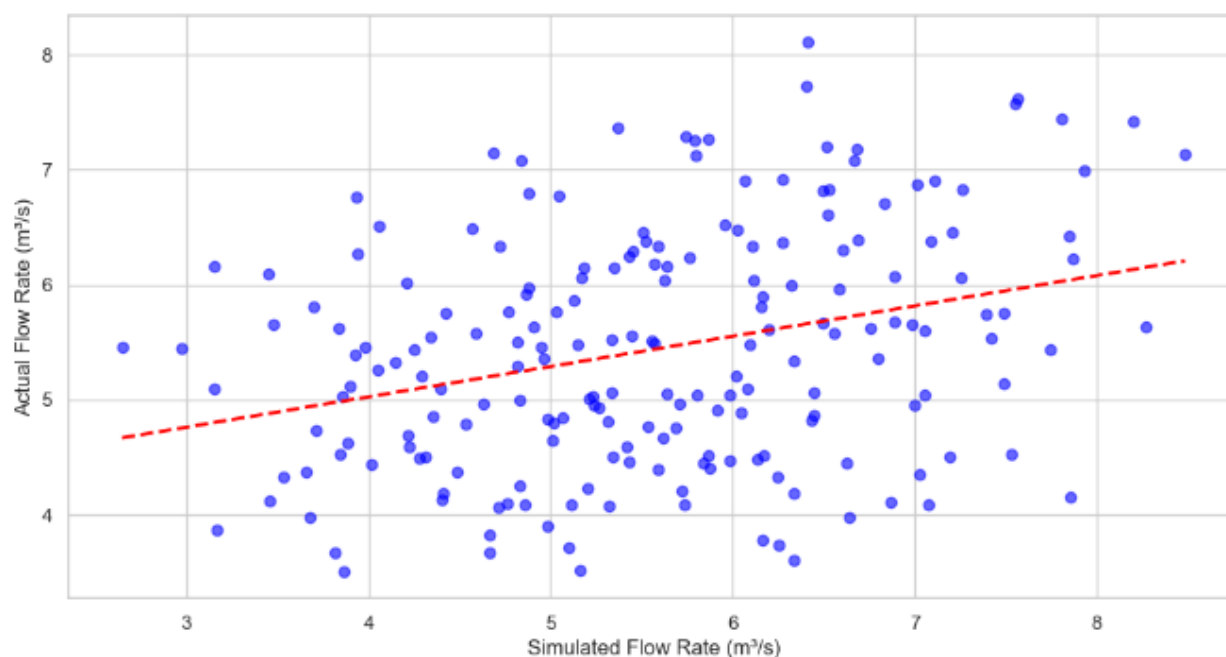


Figure 4. Actual Flow Rate against Simulated Flow Rate.

2.2.11. Model Evaluation

In this study, graphical analysis techniques were used to visually evaluate the accuracy and reliability of the newly developed correlation. The main approach employed was the cross-link method, which facilitates a straight forwarded evaluation among the actual and predicted values providing clear insights into the model's performances [13]. Three widely recognized statistical metrics Average Relative Error (ARE), Average Absolute Relative Error (AARE) and Root Mean Squared Error (RMSE) were employed to support the visual analysis and quantitative analysis to evaluate the correlation accuracy. The statistics are well-regarded for their effectiveness in determining the prediction accuracy and identifying deviation between the model and actual data [14]. Together, the graphical and statistical evaluation offered a thorough validation framework, demonstrating the new correlation's advantage over existing models.

3. Results and Discussion

This section provides detailed results and a discussion of key findings.

3.1. Modelling and Simulation

The contrast between actual and simulated flow rate values is illustrated in Figure 4. The evaluation of the simulation model's ability to represent the liquid loading behavior in gas condensate wells. The scatter plot shows how the model successfully captures the general procedure. There are substantial unconventionalities between the simulated and the

predicted values, which suggest a very weak interconnection between the operational parameters. This deviation suggested that the model has difficulty accurately replicating the complex dynamic of the liquid loading in real field data conditions. The main contributing factors were oversimplified assumptions, modeled external variables and uncertainty in the input variables [15].

The model may not adequately represent some factors in the gas condensate reservoir such as reservoir heterogeneity, well dynamics and fluid characteristics due to their complexity [16]. These discoveries highlight the need for improved model calibration to exhibit a good account of real-world variability.

3.2. Machine Learning Algorithm

Comparative analysis of the actual flow rate with the predicted flow rate as shown in Figure 5 which was developed by a machine learning model that provides a critical evaluation of the model performance in liquid loading prediction in gas condensate wells. This contrast offers a reasonable insight into the developed machine learning model's predictive capacity and its robustness across varying operational conditions.

As shown in Figure 5. Strong agreement between the actual and predicted values which indicated by the figure a very closely clustering data point around the red dash line. This indicated the effectiveness in capturing the complex, nonlinear model reliable in identifying the underlying production trends, this makes the ML model more valuable for accurate prediction and optimization of good performances [17].

The machine learning model demonstrates a significant

improvement in predictive accuracy compared to the traditional methods previously evaluated which indicates deviations from actual filed data. The ML model successfully minimizes these disparities, providing greater simplification accuracy where the traditional methods fail to capture the intricate dynamic of the liquid loading [12]. This improvement highlights the potentiality of the model as a reliable tool for predicting liquid loading. The model was able to demonstrate a promising performance based on the statistical metrics,

additional validation is required to check the datasets assessed its robustness across varying well conditions. Factors such as reservoir property changes, operational parameters, environmental conditions, and equipment performance impact prediction accuracy. To confirm the reliability and the applicability of the model across different production scenarios, confirming its adaptability is crucial [18]. Enhancing the dataset and integrating real-time production data, will further improve the model's accuracy and practical effectiveness.

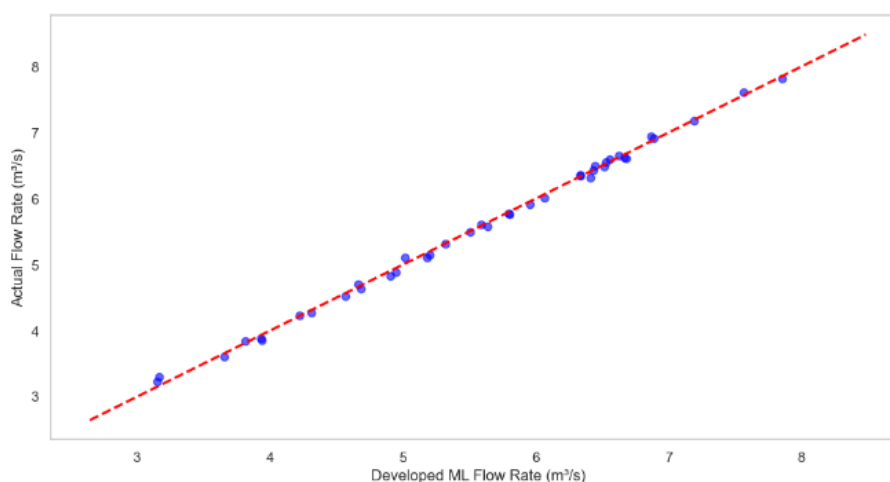


Figure 5. Actual Flow Rate against Developed Machine Learning Algorithm.

3.3. Performance Evaluation

The evaluation performance metrics summarized in Table 3 such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2) provide a very rigid assessment of the developed ML model in accurately predicting liquid loading behaviour in a gas condensate well.

Table 3. Performance Metrics of the Developed Machine Learning Algorithm.

S/N	Name	MSE	RMSE	R^2
1.	Simulation Approach	1.725	1.2389	0.8596
2.	Machine Learning Algorithm	1.561	1.1293	0.995

The developed machine learning algorithm demonstrated superior predictive performance across all metrics assessed with corresponding values of MSE and RMSE of 1.561 and 1.2389 respectively both lower than those obtained from simulation approach (1.725 and 1.2389), the model exhibits a very accurate approximation of true flow behaviors. Lower

values of MSE and RMSE are preferred as they indicate small average squared and absolute prediction errors [19].

However, the model was able to achieve an R^2 value of 0.995, suggesting it explains 99.5% variances. This reflects an exceptionally strong correlation and high predictive reliability. In comparison with the R^2 value from the simulated flow rate (0.8596), this indicates a weaker relationship and highlights the limitation of conventional modeling techniques in fully capturing complex flow dynamics [8].

These findings align with earlier research emphasizing the shortcomings of conventional simulation approaches in modeling the nonlinear and multivariable interaction characteristics of liquid loading in gas condensate wells. In contrast, the developed ML models exhibit the capability of learning from complex datasets, identifying latent patterns and generalizing effectiveness to need invisible conditions [1].

Furthermore, the model indicated lower error margins with high explanatory power positions with simplified as promising tools for operational decision-making in flow behaviors and prediction. It holds promise performance for optimizing the operational parameters in gas well performance and proactive identifying liquid loading scenarios. However, as emphasis according to Paroha [20], Additional refinement and integration of real-time production data remain essential to ensure consistent performance across diverse field operational conditions.

3.4. Future Prediction of Liquid Loading

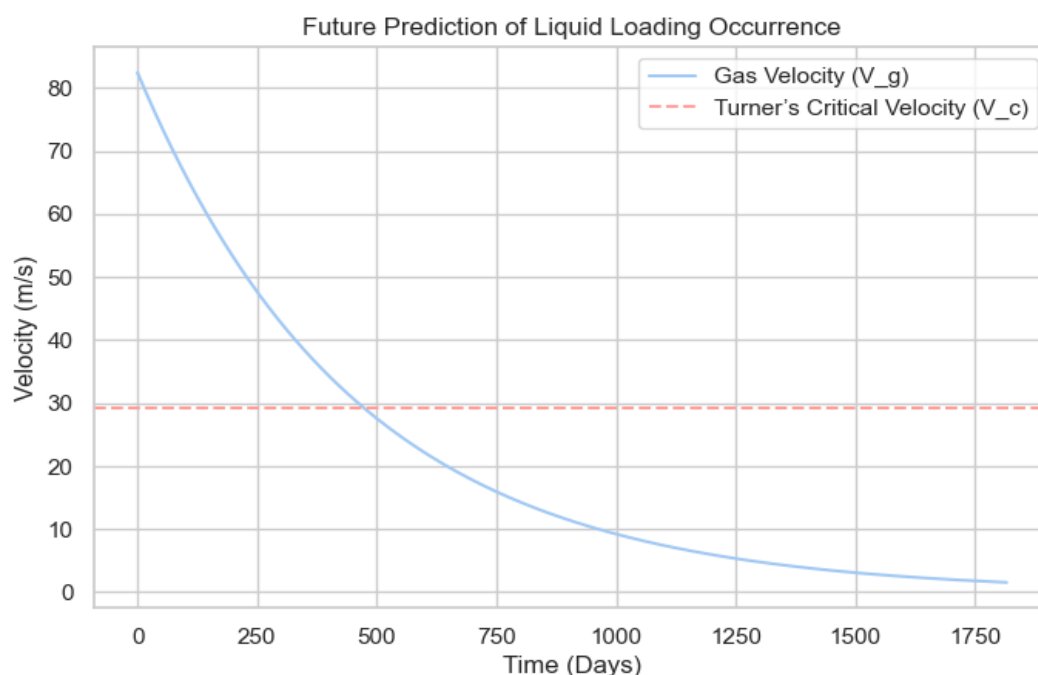


Figure 6. Liquid Loading Future Prediction.

Figure 6 shows the projected occurrence of liquid loading, providing valuable perception into the developing behaviors of gas velocity (V_g) over time and its critical, relationship with turner's velocity (V_c). A widely accepted benchmark for assessing liquid loading risk in gas well. As illustrated in the figure, gas velocity follows a typical exponential decline pattern, starting at a relatively high value and progressively decreasing over time. These findings are consistent with fundamental reservoir dynamics, where gas production rates decline over time due to reservoir pressure depletion and increased resistance to fluid flow [21].

The most notable observation in the graph is the intersection point where the gas velocity (V_g) falls below Turner's critical velocity (V_c), marking a significant beginning defined by Turner's criterion. As noted according to Liu, et al. [22] As noted according to Turner determined the gas velocity must exceed the minimum value (V_c) for effective gas flow to transport liquid droplets to the surface which is influenced by the fluid properties such as gas and liquid densities and surface tension. The loading of liquid happens when the V_g drops below Turner critical velocity, then the upward forces are no longer sufficient to overcome the gravitational setting, leading to accumulation at the wellbore.

The liquid loading is predicted to occur between 400 and 500 of the production according to the simulation results. These prediction windows are operationally significant while providing a timely opportunity to develop preventive strategies such as artificial lift methods such as plunger and gas lift

wellhead pressure optimization. These interventions are critical not only to minimize production losses but also to extend the economic viability of the well [23].

Additionally, the developed ML model's capability to simulate this transition phase with a clear data-driven approach demonstrates its readiness as a decision-making tool for the field of engineers. Unlike the conventional reactive strategies that respond only after liquid loading disrupts production, this predictive model enables proactive planning by anticipating performance shifts before critical limits are reached. However, the newly developed model offers strong theoretical insights, and further validation using real-world production data as it is essential for enhancing the accuracy and accounting for field-specific variables such as reservoir heterogeneity, completion design and transient flow dynamics [23].

3.5. Comparative Analysis

The comparative analysis of different modeling approaches, as summarized in Table 4, indicates the trade-offs between predictive accuracy, computational efficiency, and data requirements in predicting liquid loading in gas condensate wells. The new machine learning algorithm developed in this study outperforms all other evaluated models achieving an RMSE of 1.1293Mscf/d and R^2 of 0.997. The developed ML model algorithm exhibits strong performance compared to the other alternative approaches such as Hybrid Machine Learning Model (RMSE =2.8639Mscf/d, R^2 =0.9778) and Feed

Forward Neural Network ($R^2=0.9833$), shows the model effectiveness in dealing with liquid loading dynamic with relatively streamline architecture and moderates data demand [6, 23].

However, each of the models demonstrated its own strengths and weaknesses. The developed IAOM model is valued for its integration with real-time diagnostics but re-

quires intensive calibration and substantial data preparation [21]. XGBoost exhibited effective identification of onset liquid loading but highly depends on quality date training availability [22]. While the Feed-Froward Neural Network and Hybrid model prove strong predictive capabilities they are limited by high computational cost and the need for large datasets.

Table 4. Comparative Analysis of Different Methods of Predictions against Developed New Algorithm for Machine Learning.

Model Type	Strengths	Limitations	References
IAOM Framework	Real-time diagnostics, integration with digital platforms	Requires accurate calibration, data integration efforts	[24]
XGBoost Algorithm	High accuracy in predicting the onset of liquid loading	Performance dependent on training data quality	[25]
Feed-Forward Neural Network	Excellent predictive capability ($R^2 = 0.9833$)	High computational resource requirements	[26]
Hybrid Machine Learning Models	High accuracy (RMSE = 2.8639Mscf/d, $R^2 = 0.9778$)	Complex model development, need for large datasets	[6]
Developed a new machine-learning algorithm	Very high accuracy (RMSE = 1.1293Mscf/d, $R^2 = 0.997$)	Performance depends on the quality of data	and the model's simplicity

3.6. Practical Implication

The new ML algorithm developed for predicting liquid loading in gas condensate well proves a significant practical benefit by improving the production efficiency while supporting proactive operational decision-making, with strong predictive accuracy (RMSE=1.1293Mscf/d, $R^2 = 0.997$) with a minimal computational demand, the model well-matched for real-time monitoring and can be seamlessly integrated into existing digital oilfield infrastructures. It exhibits a straightforward design, coupled with performance. Offers a clear advantage over more complex and data-intensive alternatives [27]. Furthermore, its versatile adaptability makes it suitable for diverse field conditions for valuable managing mature well and marginal fields. Although the model accuracy highly depends on quality data input, advancement in modern sensor technologies is increasingly addressing these challenges by making the algorithm practical and scalable for field implementation [28].

4. Conclusion

In conclusion, this study highlights the effectiveness and practical significance of the newly developed machine learning algorithm for predicting liquid loading in gas condensate well through the flowing key findings.

- 1) *Summary of findings*; The newly developed machine learning algorithm was able to achieve a lower Mean Squared Error (MSE) of 1.561 and Root Mean Square Error (RMSE) of 1.1293, surpassing the actual flow rate baseline (MSE=1.725, RSME = 1.2389) this indicating more precise flow behaviors prediction. The model produced an R^2 value of 0.995, demonstrating that it explains 99.5% of the variance in actual flow rate data. This is significantly higher than the 0.8596 R^2 achieved using traditional or baseline methods. The newly developed machine learning algorithm outperforms previous models by achieving the highest accuracy (RMSE = 1.1293Mscf/d, $R^2 = 0.997$) with a simpler architecture. Despite high data quality requirements, this model was able to balance the performances and usability better than more complex hybrid models, which required larger datasets and a more intricate development process.
- 2) *Implication*; The model exhibited efficient integration with low computational demands and streamlined design to work perfectly into existing digital oil field systems, which makes it a practical and scalable tool for real-time application. They demonstrate strong accuracy and reliability that make them suitable for field deployment, supporting production prediction in oil wells. Also indicated performance optimization and early liquid loading detection that will enhance well management.
- 3) *Limitations*; Data quality issues such as missing values,

noise, or class imbalance introduce bias and require extensive preprocessing, increasing implementation complexity. Computational constraints restrict the use of resource-intensive algorithms, particularly in real-time systems with limited processing capabilities. Additionally, models trained on historical data often fail to adapt to new or evolving datasets, necessitating frequent retraining. Finally, domain-specific challenges, such as the dynamic conditions in oil and gas production, frequently require hybrid physics-informed ML approaches that are more complex to develop and validate than standard data-driven methods. Other limitations may include model dependency on high-quality data, risk of over fitting, generalization to other wells/reservoirs.

- 4) *Future work*; future research should focus on developing an intelligent decision support system that not only detects liquid loading but also recommends mitigation strategies in real time. This system would integrate predictive analytics with prescriptive optimization to transform early warnings into actionable solutions.

5. Recommendations

The following were discovered from the research that

- 1) Oil and gas companies should integrate real-time data acquisition systems into well monitoring.
- 2) Machine learning models should be incorporated into existing well management frameworks to improve prediction accuracy.
- 3) Operators should use predictive insights to adjust well-head pressure, gas lift systems, and optimize production strategies before liquid loading occurs.
- 4) Advanced sensor-based monitoring should be deployed to ensure continuous data collection for better decision-making.

Abbreviations

ML	Machine Learning
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
R ²	Coefficient of Determination
STB	Stock Tank Barrel
FFNN	Feed Forward Neural Networks

Acknowledgments

The authors express their sincere appreciation to the Petroleum Engineering Department of Abubakar Tafawa Balewa University Bauchi, for their unwavering support and provision of essential resources throughout this research work. Special acknowledgment to the faculty staff for their valuable guidance and encouragement which played a significant role in the successful completion of this research work.

Author Contributions

Ahmad Muhammad Salisu: Conceptualization, Methodology, Software, Data assembling, Writing- Original draft preparation.

Ibrahim Ayuba: Supervision, Visualization, Investigation.

Abdulrahman Abdulrasheed: Software, Validation.

Abubakar Usman: Writing- Reviewing and Editing.

Data Availability Statement

The data is available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] M. J. Khodabakhshi and M. Bijani, "Predicting scale deposition in oil reservoirs using machine learning optimization algorithms," *Results in Engineering*, vol. 22, p. 102263, 2024.
- [2] Y. Chen *et al.*, "Study of the Optimization of Pressurization Timing and Parameters for Enhanced Well Production Based on an Integrated Wellbore-Gas Reservoir Coupling Dynamic Analysis Method for Shale Gas Wells," *Processes*, vol. 13, no. 4, p. 1058, 2025.
- [3] P. Chen, Y. Chen, C. Yang, Y. Xu, and G. Feng, "Gas well production optimization: Classifying liquid loading severity in shale gas wells using semi-supervised learning," *Gas Science and Engineering*, vol. 128, p. 205394, 2024.
- [4] H. Xiao, S. He, M. Chen, C. Liu, Q. Zhang, and R. Zhang, "Two-Phase Production Performance of Multistage Fractured Horizontal Wells in Shale Gas Reservoir," *Processes*, vol. 13, no. 2, p. 563, 2025.
- [5] S. A. Garini, A. M. Shiddiqi, W. Utama, and A. N. F. Insani, "Filling-well: An effective technique to handle incomplete well-log data for lithology classification using machine learning algorithms," *MethodsX*, vol. 14, p. 103127, 2025.
- [6] S. Deng *et al.*, "A hybrid machine learning optimization algorithm for multivariable pore pressure prediction," *Petroleum Science*, vol. 21, no. 1, pp. 535-550, 2024.
- [7] M. Rabiei, K. Venugopal, K. Balaji, C. Abdelhamid, and A. Latrach, "Data Analytics, Machine Learning, and Artificial Intelligence in Unconventional Resources," in *Unconventional Resources*: CRC Press, 2025, pp. 565-626.
- [8] A. A. Ewees, M. A. Al-qaness, H. V. Thanh, A. M. AlRassas, and M. A. Elaziz, "Optimized neural networks for efficient modeling of crude oil production," *Knowledge and Information Systems*, pp. 1-22, 2025.

- [9] L. M. Pirstill, Q. Yue, F. R. H., and C. and Kharangate, "Statistical and machine learning applied to the universal consolidated database to predict heat transfer coefficient in flow condensation," *Numerical Heat Transfer, Part B: Fundamentals*, vol. 85, no. 11, pp. 1599-1626, 2024/11/01 2024, <https://doi.org/10.1080/10407790.2023.2269307>
- [10] R. Ellahi, N. Khalid, A. Zeeshan, S. M. Sait, and M. Khan, "Heat transfer flow of non-Newtonian eyring-powell fluid with mixed convection heterogeneous and homogeneous reactions using linear regression based machine learning approach," *Machine Learning*, vol. 114, no. 6, p. 147, 2025.
- [11] J. Mugisha, A. Shchipanov, and A. M. Øverland, "A New Interpretation Approach to Detect Induced Fracture Opening with Pressure Transient Analysis of Step-Rate Tests," *Geoenery Science and Engineering*, p. 213759, 2025.
- [12] Z. Fan, X. Liu, Z. Wang, P. Liu, and Y. Wang, "A novel ensemble machine learning model for oil production prediction with two-stage data preprocessing," *Processes*, vol. 12, no. 3, p. 587, 2024.
- [13] T. Ahmed, *Reservoir engineering handbook*. Gulf professional publishing, 2018.
- [14] Y. A. Kaplan, G. G. Tolun, and A. G. Kaplan, "A new approach for predicting solar radiation based on a pattern search algorithm," *Theoretical and Applied Climatology*, vol. 156, no. 1, p. 31, 2025.
- [15] W. Bai, S. Cheng, Y. Wang, D. Cai, X. Guo, and Q. Guo, "A transient production prediction method for tight condensate gas wells with multiphase flow," *Petroleum Exploration and Development*, vol. 51, no. 1, pp. 172-179, 2024/02/01/ 2024, [https://doi.org/10.1016/S1876-3804\(24\)60014-5](https://doi.org/10.1016/S1876-3804(24)60014-5)
- [16] A. Elyasa, A. Hassan, M. Mahmoud, R. Gajbiye, A. El-Husseiny, and I. S. Abu-Mahfouz, "Mitigating Liquid Loading in Gas Wells Using Thermochemical Fluid Injection: An Experimental and Simulation Study," *ACS omega*, vol. 9, no. 28, pp. 31081-31092, 2024.
- [17] N. B. Shaik, K. Jongkittinarukorn, and K. Bingi, "XGBoost based enhanced predictive model for handling missing input parameters: A case study on gas turbine," *Case Studies in Chemical and Environmental Engineering*, vol. 10, p. 100775, 2024.
- [18] D. Li, S. You, Q. Liao, M. Sheng, and S. Tian, "Prediction of shale gas production by hydraulic fracturing in changning area using machine learning algorithms," *Transport in Porous Media*, vol. 149, no. 1, pp. 373-388, 2023.
- [19] Y. Ahmed, K. R. Dutta, S. N. C. Nepu, M. Prima, H. AlMohamadi, and P. Akhtar, "Optimizing photocatalytic dye degradation: A machine learning and metaheuristic approach for predicting methylene blue in contaminated water," *Results in Engineering*, vol. 25, p. 103538, 2025.
- [20] A. D. Paroha, "Real-Time Monitoring of Oilfield Operations with Deep Neural Networks," in *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, 2024: IEEE, pp. 176-181.
- [21] J. Jiang, K. Li, J. Du, Z. Chen, Y. Liu, and Y. Liu, "Prediction system for water-producing gas wells using edge intelligence," *Expert Systems with Applications*, vol. 247, p. 123303, 2024.
- [22] H. Liu *et al.*, "A modified comprehensive prediction model for wellbore temperature-pressure field and liquid loading of gas wells," *Geoenery Science and Engineering*, vol. 232, p. 212452, 2024.
- [23] H. Zhang *et al.*, "An Evaluation of the Applicability of the Steady-State Productivity Approach for Horizontal Wells in Low-Permeability Heterogeneous Gas Reservoirs," *Processes*, vol. 13, no. 1, p. 173, 2025.
- [24] F. Messina, D. Rosaci, and G. M. Sarnè, "A Neural-Symbolic Approach to Extract Trust Patterns in IoT Scenarios," *Future Internet*, vol. 17, no. 3, p. 116, 2025.
- [25] G. Grekousis, "Geographical-XGBoost: a new ensemble model for spatially local regression based on gradient-boosted trees," *Journal of Geographical Systems*, pp. 1-27, 2025.
- [26] A. Nugroho and A. Choiruddin, "Forecasting Pipelines Operating Pressure: A Proactive Approach to Prevent Oil Congealing Using ARIMA, FFNN, and Hybrid Models," in *2024 IEEE International Symposium on Consumer Technology (ISCT)*, 2024: IEEE, pp. 457-463.
- [27] V. P. C. R. Udumula, R. Ancha, M. K. Ramayanam, P. V. S. M. Teja, and V. Rachpudi, "An Enhanced Real-Time Object Detection Method using Liquid Neural Network and Echo State Network Architecture," in *2024 International Conference on Inventive Computation Technologies (ICICT)*, 2024: IEEE, pp. 669-677.
- [28] F. Wang, Y. Zhang, Y. Xu, and Q. Zheng, "Lightweight Real-Time Network for Multiphase Flow Patterns Identification Based on Upward Inclined Pipeline Pressure Data," *Flow Measurement and Instrumentation*, p. 102840, 2025.