



## **A MACHINE LEARNING-BASED OPTIMIZATION FRAMEWORK FOR ESTIMATING GAS INJECTION AND ENHANCING OIL RECOVERY IN PETROLEUM RESERVOIRS**

**Alaidany, A. A.<sup>1</sup>, Altimary, M.<sup>2</sup>, Attoof Mohamed-Saeed, H.<sup>3</sup>, and Ali Abdulabbas, Z.<sup>4</sup>**

<sup>1</sup> *Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq.*

<sup>2</sup> *Department of Fuel and Energy Techniques Engineering, Shatt Al-Arab University College, Basra, Iraq.*

<sup>3</sup> *Department of Computer Science, College of Science, Mustansiriyah University, Iraq.*

<sup>4</sup> *College of Pharmacy, Al-Zahraa University for Women, Iraq.*

<sup>1</sup>[ali.ahmmad@sa-uc.edu.iq](mailto:ali.ahmmad@sa-uc.edu.iq)

### **ABSTRACT**

**Purpose:** This study presents an integrated data-driven framework for optimising gas injection strategies in enhanced oil recovery processes.

**Design/Methodology/Approach:** The proposed framework uses Long Short-Term Memory (LSTM) to model nonlinear temporal dependencies. The key operational and reservoir variables, such as gas injection rate (GIR), bottom-hole pressure, separator pressure, reservoir temperature, tubing inner diameter, gas-oil ratio (GOR), and gas composition, are considered. In the proposed method, after preprocessing, feature selection is done using the Sequential Forward Selection (SFS) method. Then, the Whale Optimisation Algorithm (WOA) was employed to optimise injection strategies and tune LSTM hyperparameters. The objective is to maximise the net present value (NPV) subject to operational constraints.

**Research Limitation:** Uncertainties and changing reservoir conditions may limit the generalisability of the proposed framework without further real-time adaptation.

**Findings:** The results demonstrate that the hybrid WOA–LSTM framework outperforms LSTM and GRU models in both prediction accuracy and economic evaluation. In the multivariate scenario, the model's RMSE is 2.22, MAE is 1.09, accuracy is 97.20%, and NPV is \$27.42 million. The results confirm the effectiveness of integrating metaheuristic optimisation and deep learning to enhance production forecasting and decision-making.

**Practical Implication:** It can enable oil field operators to improve the production efficiency and maximise economic returns while respecting the operational constraints.

**Social Implication:** Optimising gas injection resources reduces waste and enhances energy efficiency. In air pollution, the proposed model reduces hydrocarbon production and improves air quality.

**Originality/Value:** This study introduces a novel hybrid framework that combines LSTM-based forecasting with WOA for prediction. The model offers a powerful solution for complex reservoir management problems.

**Keywords:** *Deep learning. energy efficiency. gas injection. oil reservoirs. optimisation*



## INTRODUCTION

The transition from continuous growth to natural decline, along with the world's energy demand for output from several oil fields, has increased the need for solutions that effectively enhance recovery (Malozyomov et al., 2023; Agwu et al., 2024).

Gas injection is one of the approved EOR methods that could help improve ultimate recovery. It can be done all at once or in cycles. According to Moelyono et al. (2025), it helps maintain the tank's internal pressure, reducing the ratio movement and making fluid flow easier. To capitalise on this potential and achieve economically beneficial injection rates and patterns in the long run, a well-planned strategy within reasonable limitations is necessary. costs, surface geological uncertainty, changing reservoir dynamics, and gas supply limits are among the other factors that worsen the problem, according to Krishnan et al. (2024).

Classical methods based on scenario-based simulation or static settings, especially in the presence of incomplete data and operational shocks, often produce outliers and risky responses or incur high computational costs that prevent rapid updates (Ali et al., 2025; Temizel et al., 2020; Kalogirou, 2003).

Recent advances in data-driven models and machine learning (ML) have provided new perspectives for bridging “prediction” and “control” in intelligent field management (Karimi et al., 2025). Time-oriented deep learning models, especially LSTMs, have demonstrated their effectiveness in extracting long-term dependencies from operational time series, such as bottomhole pressure, production/injection rates, and composite indices of gas-to-oil ratio (GOR) or water-to-oil ratio (WOR), in process domains. This capability provides a basis for surrogate models that can predict reservoir response over short to medium time horizons with appropriate accuracy and speed, rather than using heavy simulators.

In addition, metaheuristic algorithms such as the Humpback WOA ( Mirjalili & Lewis, 2016) provide a derivative-free framework for efficient search in continuous/discrete decision space and reduce the risk of getting stuck in local optima by balancing exploration and exploitation (Alaidany et al., 2023; Zhou et al., 2021). The combination of these two paradigms, namely time-based prediction and derivative-free optimisation, provides a suitable basis for adaptive allocation of injected gas at the well/model level, under technical and economic constraints.

Despite the growing body of research in production optimization, significant gaps remain: (i) many approaches are limited to static settings and do not ensure decision robustness in the face of changing flow regimes, such as sudden changes in GOR or water inflow; (ii) a significant portion of studies lack a rolling-horizon update loop for field deployment due to their reliance on expensive simulators; and (iii) systematic integration of the economic objective function of NPV with strict safety/operational constraints (formation fracture pressure, surface injection/treatment ceiling, environmental constraints) into a scalable data-driven framework



is less well-reported (Rana et al., 2020) Furthermore, the quality of field data—noise, lack of uniform sampling, and changing well configurations—requires mechanisms to reduce model sensitivity and continuous calibration (Waqar et al., 2023; Feng et al., 2022).

Aiming to fill these gaps, this research presents an optimisation prediction framework for gas injection optimisation, whose key elements are:

- LSTM-based surrogate model to predict key reservoir responses in horizons  $t+1 \dots t+H$  with block time validation and information leakage prevention settings;
- Metaheuristic optimisation with WOA for two decision layers: (a) tuning LSTM hyperparameters (e.g., memory size, learning rate, time window) to maximise prediction accuracy and stability, and (b) allocating GIR share among wells under operational/safety constraints;
- Explicit economic-constraint formulation in which the risk-adjusted NPV is optimised as the main objective function and violations of hard constraints are controlled through a safety layer and structured penalties;

In this framework, the feature set includes time series of bottomhole pressure, oil/gas/water production rates, injection rate and pressure, and composite indices (GOR/WOR), enriched with noise-resistant preprocessing (robust averaging, robust statistical outlier removal, and time-based normalisation) and an operational calendar. The LSTM model, with its multilayer architecture and memory gates, predicts future reservoir responses and estimates the uncertainty matrix (via Monte Carlo ensemble/Dropout) so that the optimiser can make a risk-averse decision in the next step. WOA then optimises the injection rate/pattern for each well using proportional functions that maximise NPV subject to fracture pressure, surface capacity ceiling, and available gas limit constraints, achieving a balance among economic value, gas consumption, and constraint compliance. This procedure is periodically re-executed in sliding time windows to keep the injection policy robust to data noise and changing reservoir conditions. The innovation of this work can be summarised as follows:

- Provide an LSTM with uncertainty estimation to predict reservoir response under gas injection policies; Provide a two-level WOA-based optimisation that simultaneously and derivatively tunes both the hyperparameters of the forecast model and the operational gas allocation.
- Integrate the economic objective function NPV and safety/operational constraints in an explicit formulation that is transferable to industrial environments;

The rest of the paper is structured as follows: Section 2 presents a review of the relevant literature and the place of data-driven/heuristic methods in EOR; Section 3 describes the data, preprocessing, mathematical formulation of the problem, and the architecture of LSTM and WOA; Section 4 reports the evaluation protocol, sensitivity scenarios, and comparative results; and Section 5 discusses the conclusion and future directions.



## **LITERATURE REVIEW**

Given the uncertain and dynamic nature of oil reservoirs affected by geological heterogeneity, pressure drop, GOR/WOR variations and multiphase mixing, production control and development scenario design require adaptive strategies that go beyond the principles of solution drift, water drift, gas drift and gravity drift (Dietrich & Uniyal, 2020) and are aligned with diverse gas injection methods, miscible/immiscible and continuous patterns, WAG and cyclic; with CO<sub>2</sub>, N<sub>2</sub> and natural/recycled gas (Zhu & Furui, 2006).

Industrial evidence shows that the choice of gas type, boundary pressures, and operating pattern determine displacement efficiency and the path of GOR/WOR evolution; therefore, decision-making should be guided by safety/operational constraints and economic objectives. From a modelling perspective, the range of approaches extends from analytical (dependent on simplifying assumptions) to black-box empirical, federated (semi-empirical), and finally data-driven/ML, where deep temporal models such as LSTMs, trained on field data, can predict reservoir response and support real-time injection optimisation. Accordingly, the main gap in the literature is the need for frameworks that link time series forecasting with independent derivative optimisation and industrial integration to achieve optimal and sustainable gas allocation under uncertainty; this is the direction that will be followed in the following review of relevant research.

In Zeinolabedini et al. (2025), this study focuses on the optimisation of continuous gas injection (CGL) using artificial intelligence: four synthetic multilayer perceptron (MLP), radial basis function (RBF), general regression neural network (GRNN), and cascade forward neural network (CFNN) were developed to predict bottomhole pressure and production rate, and then multiple metaheuristics were tested to determine the optimal injection rate. However, the sole reliance on simulation data (without field validation), the neglect of fluid composition dynamics, the simplification of project economics, and the lack of consideration of deep learning methods are key limitations of the research.

Movahed et al. (2024) developed an artificial neural network (ANN) model, and production logging tool (PLT) and well-test separation data were extracted. First, the neural network model was trained and tested using synthetic data generated by the PROSPER1 software. The model was then evaluated using a set of test points collected from the production logging tool (PLT) logs. The results of this study showed that the ANN model has a high ability to accurately predict bottomhole pressure and well fluid flow rate.

Ramos et al. (2024) present a closed-loop, iterative gas-injection optimisation workflow for each well, which has been implemented across more than 1300 wells owned by ExxonMobil in the Permian Basin. In this method, the optimal GIR is determined by analysing the relationship between GIR and bottomhole pressure, thereby maximising the pressure difference and optimising production. An ML model is developed to provide suggestions for optimising the GIR.

ISSN: 2408-7920

Copyright © African Journal of Applied Research

Arca Academic



Mahdi (2023) presents a comprehensive review of the field of artificial lift method selection, utilising innovative ML techniques to optimise the selection process and improve accuracy. This research demonstrates the profound impact of ML applications in artificial lift method selection and uses both supervised learning and clustering methods.

Olszak et al. (2022) present the results of a field experiment in which a combination of physical methods and artificial intelligence (AI) technology is used to automatically optimise piston lifting operations. In this study, a fully automated management system for adjusting piston operating optimum points based on AI algorithms based on physical models is implemented in a field experiment.

Katterbauer et al. (2022) present a data-driven, physics-based framework for optimising aqueous hydrogen production from depleted/flooded reservoirs using thermal EOR and in-tank combustion. Using high-temperature wireless sensors and hydrogen temperature/flow rate data, oxygen/air injection is adjusted in real time to maximise combustion stability and production efficiency; the evaluation is performed on a multi-well synthetic model. The results suggest that integrating smart sensors and data-driven optimisation can improve the efficiency and economics of heat recovery and serve as a supporting technology for the transition to green energy in the oil and gas industry.

Abdollahi et al. (2021) present an integrated model of oil production from reservoir to surface; wells are simulated with PROSPER, reservoir with REVEAL, and surface facilities with GAP, and RESOLVE integrates them all. At the same time, an artificial neural network (ANN) based on Box–Behnken experimental design is used to predict performance, achieving AARD $\approx$ 3% for training and 2% for testing. The results show that bottomhole pressure and separation pressure have opposite behaviour with respect to the cumulative oil production target, and the effect of bottomhole pressure is more complex and pronounced.

A review of the research background shows that most studies have used ML, especially ANN, along with metaheuristics such as PSO, GWO, and GA, for gas injection and artificial lift problems. Although previous work has high accuracy, it usually relies on only simulated data. The limitation of flexibility in response to reservoir dynamics leads to weak learning of long-term time dependencies. Moreover, optimisation has often been applied to static parameters and is not applied in real field conditions. In response, this study proposes a combined approach that uses an LSTM to learn long-term temporal relationships among important parameters and a WOA for efficient search in the response space. As a result, the proposed method provides an integrated data-driven framework that simultaneously trains and tunes the prediction model and optimises the injection rate in both field constraints and uncertainties. Thus, compared with previous studies, the proposed model demonstrates superior prediction accuracy, stability, adaptability to reservoir dynamics, and readiness for industrial deployment.



## **METHODS**

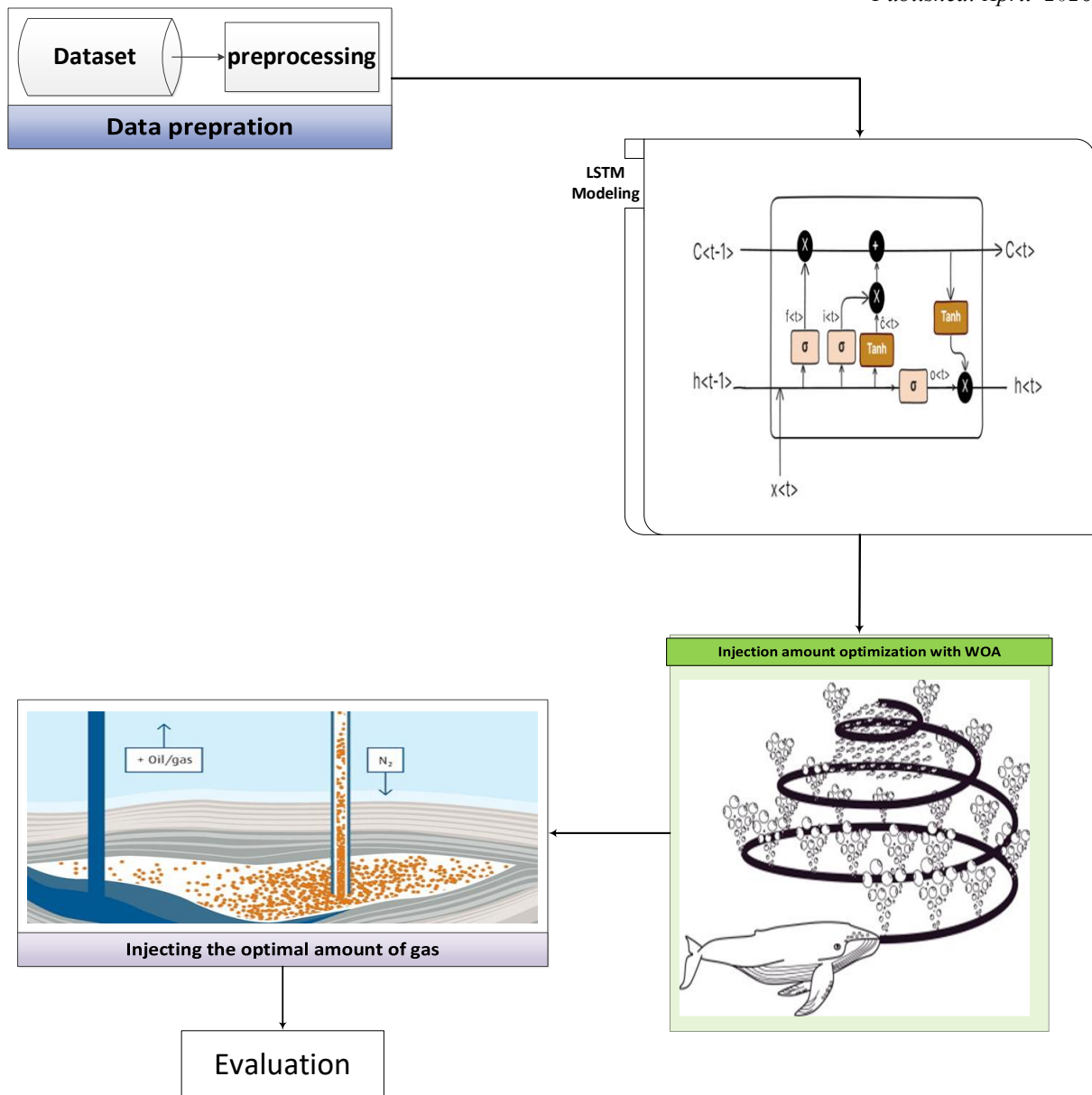
This research employs a computational, data-driven approach to develop an integrated framework for predicting and optimising gas injection performance in oil reservoirs. In this method, preprocessing of operational data and reservoir parameters is performed initially through data cleaning, removal of invalid entries, and Min-Max normalisation. So, the LSTM was used to create the prediction model. LSTM models the reservoir's changing behaviour by looking at how variables interact with each other, both linearly and nonlinearly, and how those interactions change over time.

The SFS method was also used to simplify the model by identifying the most important features. The WOA was then used to improve gas injection methods and to adjust the model's hyperparameters to maximise the net present value. Finally, it may be able to make very accurate production projections by combining these methodologies in a new way. This would provide strong evidence for the best way to manage gas injection in oil reservoirs. Figure 1 shows the general block diagram of the proposed process of modelling and optimising gas injection.

In addition, one of the important steps in the data preparation process was normalising numerical values. By converting the values into a specific range (usually between 0 and 1), this technique causes all features to be on the same scale and prevents the negative impact of the difference in the dimensions of the values on the model learning process. This is especially important in neural networks such as LSTM, which are sensitive to the size and distribution of the input values. In this method, each data value is converted to a new value in the range 0-1. This is done using the Min-Max method, whose formula is given as follows (Huang, Li, & Xie, 2015).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where  $X_{norm}$  is the normalized result. From normalization, X is the initial value before normalization. Here  $X_{max}$  and  $X_{min}$  represent the maximum and minimum values of each feature, respectively (Karimi, Harouni, Jazi, Nasr, & Azizi, 2022).



*Figure 1: Block diagram of the proposed method*



### **Modelling with LSTM**

In the first stage, the focus is on training the LSTM (Abolghasemi, Emadi, & Karimi, 2024). As a powerful deep learning tool, this model can analyse temporal dependencies and hidden trends in reservoir parameters. Historical data recorded from reservoir performance are used to train the model. These data include variables such as GIR, reservoir pressure, daily oil production rate, and other dynamic system characteristics. In this process, data is entered into the model as time sequences, and LSTM learns the relationship patterns between inputs and outputs through successive iterations.

The goal is for the model to predict outputs, such as oil production rates and economic indicators (e.g., NPV), with high accuracy across different injection rates. Due to the LSTM's recursive structure and memory retention, this model is well able to identify the late effects and long-term dependencies of gas injection on reservoir performance. The result of this step is a model that can serve as a reliable predictor in the optimisation process.

• *Model Input:*

According to the model structure proposed in this study, a set of key operational and thermodynamic variables is introduced as inputs to the LSTM model to accurately simulate the behaviour of the reservoir and production system. These inputs are shown in Table 1. A combination of input parameters offers a highly detailed representation of subsurface conditions, injected gas properties, and operational specifications. Such a comprehensive dataset is critical for effective LSTM training and accurately predicting system dynamics.

*Table 1: Model Inputs*

<b>Input (Feature)</b>	<b>Effect in the Model</b>	<b>Data Note / Pre-processing</b>
GIR	Direct control of reservoir pressure and displacement front; primary driver of the optimisation policy	Min–Max normalisation or standardisation; time alignment with targets/outputs
Bottom-hole pressure (BHP)	Indicator of flow potential and energy loss; sensitive to pattern/rate changes	Outlier removal; short moving-average filtering; use as lag/lead features
Separator pressure	Governs surface gas–liquid equilibrium and produced GOR	Periodic stabilisation of operational shifts; do not one-hot encode—keep continuous
Reservoir temperature	Affects viscosity and miscibility; slow but influential variability	Typically, quasi-stationary/slow-varying; can be included as a contextual variable
Tubing inner diameter (ID)	Influences pressure drop and flow rate; hydraulic conduit	If time-invariant, include as a static (time-independent) feature
Gas–Oil Ratio (GOR)	Proxy for volatility/pressure depletion; early warning for gas breakthrough	Separate instantaneous vs cumulative GOR; apply lagging and de-noising



Gas composition: methane/ethane / propane	Determines miscibility and EOR efficiency; impacts PVT behaviour	If missing, impute/extrapolate using reference proportions or lab/PVT data
---	--	--

The proposed architecture in Table 2 starts with an LSTM layer that converts the feature sequence ( $T \times F$ ) into a latent representation ( $T \times H$ ); with `return_sequences=True`,  $H$  sample size = 64, tanh inner activator, and sigmoid gates to model the short/long term dependencies and delay of the injection rate/pressure effect stably and to provide a substrate for deeper or Attention layers. To control overfitting, `Dropout=0.2` is applied after the recurrent part to reduce unit correlations and improve generalizability. Then a “temporal aggregation” step (global mean/max or the final Flatten step) compresses the latent sequence into a vector ( $H$ ) to provide a stable, interpretable profile for the output. Finally, a single-neuron Dense layer with linear activation maps the feature vector to a continuous target variable—such as production rate or NPV—so that the prediction range is not limited and the output can be directly used with economic/operational criteria.

*Table 2: Architecture and components of the LSTM model for gas injection optimisation*

Component	Role	Input / Output	Key Settings	Design Notes / Rationale
LSTM layer (first)	Extract short/long-term temporal dependencies	Input: feature sequence ( $T \times F$ ) → Output: latent sequence ( $T \times H$ )	<code>return_sequences=True</code> ; units ( $H=64$ ) (example); internal activations: tanh (cell), sigmoid (gates)	Stabilises gradients; models delayed effects of injection/pressure; foundation for deeper stacks or Attention
Dropout	Regularisation to reduce overfitting	Input: ( $T \times H$ ) → Output: ( $T \times H$ )	Dropout rate = 0.2	Lowers inter-neuron co-adaptation; improves generalisation; recurrent dropout can be added if needed
Temporal aggregation	Compress sequence to a feature vector	Input: ( $T \times H$ ) → Output: ( $H$ )	Global Average/Max Pooling or Flatten of the last step	Choose based on the stability and interpretability of downstream regression
Output Dense	Map to target quantity (regression)	Input: ( $H$ ) → Output: (1)	units=1, activation linear	Unbounded output range; suitable for oil rate / NPV predictions



### *Model Configuration:*

Model configuration is a key step for LSTM training. In this step, three main components are defined: the cost function, the optimiser, and the performance evaluation criteria. The correct selection of these components directly affects the performance of the final model.

In the proposed LSTM-based model for oil reservoir performance prediction, the mean square error (MSE) is selected as the cost function. MSE calculates the squared difference between the actual and predicted values, then averages them. Due to its high sensitivity to large errors, MSE is well-suited to problems with continuous outputs, such as production rate or NPV.

To optimise LSTM weights, the Adam algorithm has been used. Adam combines the advantages of two widely used optimisers, namely RMSProp and SGD, with acceleration. Adam adjusts the learning rate and gradient memory on its own to speed up the model's convergence and ensure stable training across intricate time sequences.

The mean absolute error and root mean squared error are two measures of how well the model performs. You may use these measurements to assess how well the model's predictions match real-world datasets across a range of situations.

### **Model Output**

The first output is the oil production rate, which, as a continuous value, predicts the reservoir production behaviour in response to gas injection conditions and plays a fundamental role in analysing the performance of different scenarios. The second output is the NPV, which aggregates the project's economic returns based on predicted production values and provides a measure of the profitability of the scenarios. Finally, the objective function for optimisation is the model's combined output, which uses the predicted values as inputs to an optimisation algorithm, such as WOA, to determine the path that maximises the GIR. This table clearly reflects the predictive and guiding role of the LSTM model across the technical and economic dimensions of the decision-making process.

### **Whale Algorithm Optimisation**

The WOA was first proposed by Mirjalili and Lewis (2016). This algorithm is inspired by the social behaviour of humpback whales. The Whale Algorithm starts with a set of random solutions. In each iteration, the search agents update their positions using three operators: bait encirclement, the bubble attack method, the network (extraction phase), and the search for bait (exploration phase). During bait encirclement, humpback whales identify the bait and surround it. The Whale Algorithm assumes that the best solution is the bait at the moment. After the best search agent is identified, the other search agents update their positions relative to the best search agent.



To design the binary version of the Whale Algorithm, some basic concepts of the Whale Algorithm are modified. Because the search is to be done in binary space, i.e., values 0 and 1, the agents' or whales' movement will change to 0 or 1. The main difference between the original WOA version and its binary version is the mechanism for updating the whales' positions. Also, the values are only 0 and 1. In BWOA, the value is updated with the probability of the whale's spiral movement. To binarise, a suitable transfer function is needed to map the whale's spiral movement values to binary values and update them. This transfer function causes the whales to move in a binary space.

### **Definition of the Objective Function**

In this research, the objective function serves as a key tool for evaluating the profitability of different gas injection scenarios and is defined based on the NPV economic indicator. The highlight of this approach is the use of the LSTM deep learning model to predict the oil production rate in different time intervals; in other words, the output of the LSTM model, i.e., the  $Q_t$  values are directly included in the objective function. After training on reservoir operational data, this model can predict future production rates for any proposed gas injection amount. This predicted production rate, along with the price per barrel of oil, operating costs, and discount rate, is then fed into the objective function and the economic value of the project is calculated as follows:

$$NPV = \sum_{t=1}^T \frac{(Q_t P - C_t)}{(1+r)^t} \quad (2)$$

In this relation,  $Q_t$  is the oil production rate at time  $t$  predicted by LSTM,  $P$  is the price per barrel of oil,  $C_t$  is the operating costs at time  $t$ ,  $r$  is the discount rate, and  $T$  is the total forecast period.

The ultimate goal is to maximise the NPV by finding the optimal GIR; this is achieved using the WOA metaheuristic. Therefore, the overall structure of this research combines a proposed data-driven model (LSTM) for technical production forecasting with an economic objective function for decision-making, which, in cooperation with the optimisation algorithm, yields the optimal strategy for managing gas injection in the reservoir.

In the optimisation process, one necessary and effective step is to determine the search scope for the decision variables. In this research, the main decision variable is the GIR entering the reservoir, which must be examined within a reasonable operational range. This range should be determined in a way that, on the one hand, reflects the actual field conditions and technical constraints, and on the other hand, provides sufficient space to find the optimal answer.

To this end, the search range is adjusted based on historical injection data, surface facility capacity, reservoir pressure, and safety constraints. For example, the GIR may be considered



in the range of 5 to 60 million cubic feet per day (MMscf/day). This range is represented as follows:

$$\text{Injection Rate} \in [Q_{\min}, Q_{\max}] \tag{3}$$

$Q_{\min}$  is the minimum possible injection rate (based on the injection threshold pressure or economic constraint) and  $Q_{\max}$  is the maximum acceptable injection rate (based on equipment capacity or rock fracture pressure). Precise determination of this range causes the optimisation algorithm, such as WOA, to search only in realistic areas and avoid testing infeasible or costly values. As a result, both the convergence speed and the final efficiency of the proposed solution are increased.

### WOA Implementation

In this framework, in the algorithm of Figure 2, first, with operational time series data (properties and oil rate), an LSTM-based predictive surrogate is trained to return the sequence of future production rates  $\hat{Q}_{1:T}$  given a time-dependent gas injection policy  $x_{1:T}$ . Then the economic objective function is defined exactly according to the relation NP

$$NPV = \sum_{t=1}^T \frac{(Q_t P - C_t)}{(1+r)^t}$$
 To maximise NPV, the WOA starts with an initial population of

injection policies within the operating bounds  $\text{Injection Rate} \in [Q_{\min}, Q_{\max}]$ ; at each iteration, each solution is updated with three standard WOA mechanisms: exploitation, exploration around a random individual, and spiral attack towards the best response. Each new policy is clipped to the allowable interval, evaluated with LSTM, and its NPV is calculated; if it is better, it replaces the previous one, and if it is better, the global best is also updated. After a certain number of iterations or a convergence criterion is met, the final output will be the best gas injection profile over horizon T and the corresponding maximum NPV.

```

# Pseudocode Of proposed method
Inputs:
X, y           # historical features and oil rate
P, C[1..T], r  # price (constant), costs, discount
Qmin, Qmax    # injection bounds
N, G          # WOA population, iterations
T             # horizon length

# 1) Train surrogate
LSTM = TrainLSTM(X, y)

# 2) Objective using the exact NPV formula
PredictProd(LSTM, x[1..T]) ->  $\hat{Q}_{1:T}$  [1..T] # roll forward with policy x
Objective(x):
  Qhat = PredictProd(LSTM, x)
  NPV = 0
  for t=1..T:

```



$$NPV = \sum_{t=1}^T \frac{(Q_t P - C_t)}{(1+r)^t}$$

return NPV # (add penalties here if needed)

**# 3) WOA optimizer**

WOA(Objective, Qmin, Qmax, N, G):

pop = [ rand(Qmin, Qmax, size=T) for i=1..N ] # time-varying policy over T

fit = [ Objective(pop[i]) for i=1..N ]

xbest = pop[argmax(fit)]; fbest = max(fit)

for g=1..G:

    a = 2 - 2\*(g-1)/(G-1)

    for i=1..N:

        r1=rand(); r2=rand(); p=rand(); l=rand(-1,1)

        A = 2\*a\*r1 - a

        Cc = 2\*r2

        Xi = pop[i]

        if p < 0.5:

            if abs(A) < 1:

                D = abs(Cc \* xbest - Xi) # encircling

                Xn = xbest - A \* D

            else:

                j = randint(1,N); Xr = pop[j] # exploration

                D = abs(Cc \* Xr - Xi)

                Xn = Xr - A \* D

        else:

            Dp = abs(xbest - Xi) # spiral attack

            b = 1.0

            Xn = Dp \* exp(b\*l) \* cos(2\*pi\*l) + xbest

        Xn = clip(Xn, Qmin, Qmax) # enforce Qmin ≤ x\_t ≤ Qmax

        fn = Objective(Xn)

        if fn > fit[i]: pop[i]=Xn; fit[i]=fn

        if fn > fbest: xbest=Xn; fbest=fn

    return xbest, fbest

**# 4) Run**

best\_policy, best\_NPV = WOA(Objective, Qmin, Qmax, N, G)

**RESULTS**

In this study, the LSTM recurrent neural network model has been used to predict oil production performance in response to different GIR. To compare and validate the proposed model's performance, a similar GRU network architecture has also been used. Three prediction structures have been defined for each model: LSTM and GRU alone, and LSTM combined



with WOA-LSTM. These structures have been evaluated in two different scenarios defined based on operational and reservoir data.

### **Evaluation Criteria**

The suitability of an algorithm or method for predicting the amount of gas injected depends on the results obtained. The following quantitative criteria will be used to evaluate these algorithms: mean square error, root mean square error, absolute error, precision, and NPV (Zeinolabedini et al., 2025; Shokir et al., 2017).

In this research, the LSTM model's output, which predicts the oil production rate at different times, directly affects the calculation of production revenue. These revenues are then entered into the NPV function, along with the fixed and variable costs of gas injection, and the final result is optimised as a numerical criterion. The higher the NPV value, the more economically desirable the gas injection scenario is considered. The database for this research comprises 10,000 simulated records related to the optimisation of gas injection in oil reservoirs. This data is designed to examine the effect GIR on oil production performance and economic criteria such as NPV. The variables in this database are divided into two categories of input variables and output variables in Table 3.

The input variables include GIR, bottomhole pressure, separator pressure, reservoir temperature, inner diameter of the production tubing, gas-to-oil ratio (GOR), and hydrocarbon compositions C1, C2, and C3. These variables are generated through software simulation and numerical modelling, and their changes are examined in the model outputs.

The output variables include oil production rate, project NPV and evaluated model errors, including MAE, MSE and RMSE. These values indicate how well the proposed model predicts the performance of the gas injection process and the extent to which optimisation has increased production and reduced operating costs. This database can be used to train and evaluate ML models, compare different optimisation algorithms, and perform economic and operational analyses related to the gas injection process. Using this data, the best GIR can be determined to maximise production and increase oilfield profitability.

*Table 3: Database specifications*

<b>Description</b>	<b>Unit</b>	<b>Variable Type</b>	<b>Variable Name</b>
GIR	MMSCFD	Input (continuous)	GIR
Bottom-hole pressure	psi	Input (continuous)	Bottom Hole Pressure
Separator pressure	psi	Input (continuous)	Separator Pressure
Reservoir temperature	°F	Input (continuous)	Reservoir Temperature
Tubing inner diameter	inch	Input (continuous)	Tubing ID
Gas–oil ratio	SCF/STB	Input (continuous)	GOR
Methane fraction in gas	%	Input (continuous)	C1 Content



Ethane fraction in gas	%	Input (continuous)	C2 Content
Propane fraction in gas	%	Input (continuous)	C3 Content
Oil production rate	STB/day	Output (continuous)	Oil Production Rate
NPV	USD (\$)	Output (continuous)	NPV
MAE	dimensionless	Output (continuous)	MAE
MSE	dimensionless	Output (continuous)	MSE
RMSE	dimensionless	Output (continuous)	RMSE

### **Settings**

In this study, a data-driven model based on LSTM in MATLAB is developed for reservoir performance prediction and economic estimates (including NPV). Historical injection/production data and fluid/reservoir properties are preprocessed (removal/replacement of missing data and normalisation ([0,1]) into training/validation/testing splits and monitored with cross-validation with a 10% share. Hyperparameter tuning is performed by stochastic search (learning rate, batch size, number of LSTM layers and neurons), and training is performed with Adam (adaptive learning rate advantage); the final configuration and summary of settings are given in Table 4.

*Table 4: Final configuration of the LSTM model*

<b>Parameter</b>	<b>Value</b>
Learning rate	0.01
Optimizer	Adam
Max iterations (epochs)	350
Batch size	80
Number of hidden units	100
Gradient threshold (clipping)	1
Execution environment	Auto
Number of LSTM layers	5
Sequence length	Longest
Data shuffling	Once
Activation function	Sigmoid
Hyperparameter tuning	Random search

In the first scenario (univariate forecasting), only the GIR variable was used as an input to the model to predict outputs such as oil production rate or bottomhole pressure. In the second scenario (multivariate forecasting), the model was fed with a set of reservoir and surface data, including bottomhole pressure, reservoir temperature, injected gas composition (methane, ethane, propane), oil production rate and other technical characteristics.

In all scenarios, the previous data values were provided to the model as a time series. For the univariate model, the best performance was achieved using 168 past data points (equivalent to



one week of operation). In contrast, in the multivariate model, using 24-hour data (96 data points for 4 input characteristics) yielded more accurate results. In each forecast run, a future 24-hour interval was estimated, including oil production rate and NPV.

Finally, the performance of different models was compared using quantitative criteria such as RMSE, MAE, MAPE, and NPV to identify the optimal model structure for forecasting and decision-making in gas injection. Combining LSTM with WOA yielded better results than the base models in most scenarios.

### **Univariate Time Series Forecasting**

In this study, the “consumption/performance” forecast was first performed as a univariate and based solely on the history of the same variable; three structures were tested: LSTM, GRU and WOA-LSTM. In the proposed univariate approach, only one key input (e.g., separator pressure, reservoir temperature, or methane percentage) was provided to the LSTM at a time to extract temporal patterns of that variable; WOA then adjusted the GIR using the model output to maximise NPV and forecast accuracy. Although this analysis lacks the comprehensiveness of multivariate analysis, it reveals the relative importance of each feature and the model's learning power with limited data. The results also showed that WOA-LSTM provides acceptable accuracy compared to the baseline models, even in univariate mode, although in multivariate scenarios it achieves significant improvements in accuracy and generalisation.

For each of the input variables tested—separator pressure, reservoir temperature, production tubing internal diameter, GOR, and methane/ethane/propane percentages in the injected gas—the same pattern is observed: the WOA-LSTM model (i.e., the LSTM with hyperparameters tuned by the WOA) consistently provides the lowest RMSE/MAE and the highest accuracy compared to the simple LSTM and GRU, and results in a higher economic NPV.

In the ranking of methods, LSTM usually comes after WOA-LSTM, and GRU shows the weakest relative performance; this trend holds across all inputs, though the magnitude of the performance gap between methods varies slightly by variable type. In each case—separator pressure (Table 5), reservoir temperature (Table 6), production tubing internal diameter (Table 7), gas dissolved in oil ratio (GOR) (Table 8), and methane percentages, Ethane and propane in gas mixture (Tables 9–11)—WOA-LSTM method systematically provides lower prediction error (lower RMSE/MAE) and higher accuracy than LSTM and GRU, while leading to higher NPV. In the ranking of methods, LSTM is usually followed by WOA-LSTM, and GRU shows the weakest relative performance; this trend is observed in all the aforementioned tables, although the efficiency gap between the methods changes slightly depending on the type of input variable.

In more detail, Tables (5) and (7) show that the operational variables close to control (separator pressure and internal diameter) with the WOA-LSTM framework provide the most consistent improvement in accuracy and economic outcome; Table (6) shows the effect of thermodynamic



inputs (reservoir temperature); and Tables (8) to (11) show the role of miscibility indices (GOR and methane/ethane/propane components) in improving economic forecasting and decision-making. The results of all tables establish the clear superiority of the WOA-LSTM forecasting-optimisation approach in the three dimensions of accuracy, stability, and economic value.

*Table 5: Input Variable: Separator Pressure*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	4.85	3.12	86.40	19.85
GRU	5.18	3.45	84.30	18.92
WOA-LSTM	3.74	2.21	91.60	22.30

*Table 6: Input Variable: Reservoir Temperature*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	5.27	3.68	85.20	18.67
GRU	5.62	3.91	83.40	17.89
WOA-LSTM	4.02	2.38	90.10	21.56

*Table 7: Input Variable: Tubing Inner Diameter*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	4.95	3.23	86.10	19.10
GRU	5.31	3.51	84.20	18.27
WOA-LSTM	3.88	2.18	91.30	22.74

*Table 8: Input Variable: Gas–Oil Ratio (GOR)*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	4.91	3.14	86.20	19.35
GRU	5.23	3.43	84.10	18.50
WOA-LSTM	3.79	2.11	91.70	22.58

*Table 9: Input Variable: Methane Fraction in Injected Gas*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	5.35	3.75	84.70	18.50
GRU	5.71	3.98	83.10	17.63
WOA-LSTM	4.08	2.49	89.80	21.24

*Table 10: Input Variable: Ethane Fraction in Injected Gas*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	5.23	3.68	85.00	18.94
GRU	5.57	3.87	83.30	18.10
WOA-LSTM	3.96	2.31	90.50	21.85



*Table 11: Input Variable: Propane Fraction in Injected Gas*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	5.07	3.48	85.60	19.28
GRU	5.39	3.71	83.90	18.33
WOA-LSTM	3.85	2.27	91.10	22.12

### **Multivariate Time Series Forecasting**

In this section, we formulate and report the feature selection problem for data-driven gas injection models with a systematic approach consistent with Q1 standards. To increase computational efficiency and improve generalizability, SFS is used. In this method, the search starts from the empty set, and at each step, the feature that provides the greatest improvement in model performance (in terms of accuracy/error reduction) is added to the selection basket. The process stops until there is no significant improvement or a predetermined threshold is reached. The advantage of SFS is its simplicity, interpretability, and low computational cost, which directly lead to reduced redundancy, reduced overfitting, and improved LSTM convergence. In the proposed framework, SFS is coupled with WOA-based hyperparameter optimisation to perform feature selection and model tuning synergistically.

Step-by-step SFS results. According to Table (12)—with two inputs “separator pressure” and “reservoir temperature”—the WOA-LSTM model outperforms the baseline models and reveals the synergy of pressure-temperature in improving the prediction of the optimal injection rate and NPV. As detailed in Table 13, incorporating the inner diameter of the production pipe as an additional feature strengthens the model’s representation of pressure drop and hydraulic behavior. This inclusion effectively reduces the prediction error compared to the two-feature baseline, with only a marginal increase in computational complexity. Furthermore, Table 14 illustrates the addition of the GOR alongside the previous variables. This enhancement enriches the thermodynamic characterisation of the fluid, noticeably minimising error margins and improving the precision of injection decision-making. Finally, as shown in Table 15, integrating methane content into the injected gas improves the accuracy of simulations of fluid miscibility and gas quality impacts. Consequently, this leads to greater prediction stability and an improved NPV.

Then, the addition of “percent ethane” in Table 16 further completes the miscibility dimension and shows that the collective effect of light components is better reflected than individual analyses; a further reduction in error and an increase in accuracy are the results of this enrichment.

The extension to “percent propane” in Table 17 as the heaviest component examined—has made the thermodynamic flow dependencies more precise and revealed more complex relationships between gas composition and reservoir response; the consequence is a further reduction in RMSE/MAE and improved injection economic policies. Finally, Table 18 with



the full set of eight features shows that the WOA-LSTM framework combined with SFS in the richest configuration records the best prediction performance and the highest NPV; This result confirms that multidimensional information (operational, thermodynamic, and hydraulic) when integrated into an integrated prediction-optimisation model, provides the greatest efficiency for field decision-making.

Methodological Summary. (i) SFS controls the problem dimension and reduces the risk of overfitting by eliminating weak/correlated features; (ii) WOA calibrates the LSTM in the stable performance region by searching for independent derivatives in the hyperparameter/policy space; and (iii) the incremental path of Tables 12 to 18 documents the role of each feature in reducing error, increasing accuracy, and improving NPV as a systematic ablation analysis. To fully comply with Q1 standards, it is recommended that the final version explicitly report and control for temporal information leakage in cross-sectional assessments, in addition to comparing means, confidence intervals, and significance tests (e.g., bootstrap or nonparametric tests).

*Table 12: Model performance with 2 input features*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	4.60	2.97	87.10	20.35
GRU	4.88	3.19	85.80	19.45
WOA-LSTM	3.73	2.20	91.80	22.48

*Table 13: Model performance with 3 input features*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	4.32	2.75	88.30	21.10
GRU	4.57	2.91	86.90	20.06
WOA-LSTM	3.41	1.98	92.90	23.31

*Table 14: Model performance with 4 input features*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	4.05	2.53	89.60	21.85
GRU	4.29	2.64	88.10	20.82
WOA-LSTM	3.10	1.75	94.00	24.19

*Table 15: Model performance with 5 input features*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	3.78	2.32	90.80	22.43
GRU	4.02	2.39	89.30	21.41
WOA-LSTM	2.85	1.55	95.10	25.02



*Table 16: Model performance with 6 input features*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	3.54	2.12	91.90	23.10
GRU	3.76	2.15	90.10	22.05
WOA-LSTM	2.61	1.38	96.00	25.87

*Table 17: Model performance with 7 input features*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	3.31	1.94	92.60	23.76
GRU	3.51	1.93	91.00	22.68
WOA-LSTM	2.38	1.22	96.70	26.65

*Table 18: Model performance with 8 input features (full set)*

<b>Method</b>	<b>RMSE</b>	<b>MAE</b>	<b>Accuracy (%)</b>	<b>NPV (million USD)</b>
LSTM	3.12	1.79	93.30	24.38
GRU	3.33	1.81	91.60	23.30
WOA-LSTM	2.22	1.09	97.20	27.42

The results of this study show that the proposed integrated framework, based on LSTM optimised with the WOA and SFS, delivers better prediction and optimisation performance than many of the methods reported in recent studies. Previous research has demonstrated that ML can effectively model the complex nonlinear relationships present in reservoir systems. In Table 19, Zhang et al. (2020) used methods such as Random Forests and ANNs to predict oil production and reservoir behaviour, showing that data-driven approaches can achieve suitable accuracy in analysing reservoir data. However, the error reported in their study (around RMSE=3.65) is higher than the value obtained in the present research (RMSE=2.22). This significant improvement can largely be attributed to the use of recurrent networks like LSTMs, which can extract temporal dependencies in production and operational data—a capability many traditional models lack.

The superiority of recurrent models in analysing reservoir time-series data was also emphasised in the study by Al Shalabi and Al Hashem (2021). They used LSTM to predict reservoir pressure and GOR, demonstrating that these models perform better compared to many common ML algorithms. The findings of the present study align with these results; however, in this research, using the WOA to tune the model’s hyperparameters significantly improved LSTM performance and reduced error. This indicates that combining deep learning models with metaheuristic algorithms can enhance the stability and prediction accuracy in complex reservoir problems.



Using optimisation methods not only improves prediction accuracy but also helps managers make better decisions about oil reservoirs. Ghorbani et al. (2022) state that using neural networks with metaheuristic algorithms such as PSO can enhance GIR and increase oil output by about 6-10%. The results of the present study are consistent with these findings and confirm the importance of using metaheuristic optimisation methods in reservoir management. However, the proposed framework in this research goes a step further by incorporating the economic metric of NPV into the decision-making process, in addition to optimising production performance. Based on the results, the optimal scenario generated an NPV of \$27.42 million, demonstrating the economic advantage of the proposed approach.

Another important result of this study is the significant impact of feature selection on improving model performance. In the research by Li et al. (2023), recurrent models like GRU and LSTM were used for multivariate prediction of reservoir behaviour, but that study did not employ systematic feature selection methods. The absence of such methods in high-dimensional data problems can lead to reduced model efficiency and increased prediction error. In contrast, this study used the SFS method to select the most important reservoir and operational parameters, including separator pressure, reservoir temperature, tubing diameter, GOR, and gas compositions. This process led to a remarkable improvement in model performance, reducing the MAE to 1.09 and increasing prediction accuracy to 97.20 percent.

Overall, comparing the results of this study with previous research indicates that the proposed WOA–LSTM–SFS framework provides a comprehensive and efficient approach for optimising gas injection in oil reservoirs. Integrating temporal modelling, feature selection, and metaheuristic optimisation into a unified data-driven framework not only increases prediction accuracy but also enables decision-making based on economic criteria. These results demonstrate that using combined ML frameworks can play a significant role in developing intelligent reservoir management systems and improving the productivity and sustainability of hydrocarbon production.



*Table 19: Literature Comparison*

No.	Article	Methods	Study Objective	Key Results	Comparison with the Present Study
1	Zhang et al. (2020),	Random Forest, ANN	Prediction of oil production and reservoir behaviour	Production RMSE: 3.65	The proposed WOA-LSTM model has higher accuracy (RMSE = 2.22) and also includes economic optimisation, which was not considered in their study.
2	Al-Shalabi & Al-Hashem (2021)	LSTM (without optimisation)	Prediction of reservoir pressure and GOR	Pressure RMSE: 3.12	The WOA- LSTM leads to improved prediction accuracy and better model stability.
3	Ghorbani et al. (2022),	PSO-ANN	Optimisation of GIR and oil production	Production increase: 6–10%	Unlike their work, the proposed framework simultaneously optimises gas injection and maximises (NPV = 27.42 million USD).
4	Li et al. (2023).	GRU, LSTM	Multivariable reservoir prediction	MAE: 1.98	The WOA-LSTM has a lower prediction error (MAE = 1.09) and demonstrates the benefit of feature selection.

## CONCLUSION

Managing gas injection into oil reservoirs often suffers from a disconnect between the predictive model and actual field decisions. To address this, the proposed framework relies on LSTM and WOA. LSTM acts as a fast surrogate model. Because it inherently captures complex, time-dependent reservoir dynamics, it can quickly test various injection scenarios without running computationally expensive numerical simulations. The SFS technique is used for selected relevant features. Once the predictions are solid, the WOA steps in to identify the optimal GIR and patterns, navigating strict operational constraints to maximise NPV.

What truly sets this workflow apart is how it ties these elements together. A significant portion of existing literature either stops at predicting reservoir behaviour or tries to optimise operational settings in isolation. In contrast, time-series forecasting, feature selection, and economic optimisation are integrated into a single, continuous pipeline. This means the system



does not just forecast what might happen. It serves as a tool for advice, assisting engineers in determining the precise amount of gas to inject to optimise profitability.

It makes sense to switch from DL-based reservoir management when considering the industrial ramifications. It reduces operational planning risk, minimises gas waste, and lowers overall operational expenditures. In an industry under constant pressure to be both efficient and sustainable, deploying a fast, reliable decision-support tool is a major operational advantage.

The course, the current methodology has its limitations. Since the framework was largely trained and tested on simulation-derived data, the next critical step is to rigorously validate it against noisy, real-world field data. The proposed method also needs to account for the inherent uncertainties of subsurface modelling, such as shifting fluid properties or unexpected flow boundaries. Moving forward, expanding this setup to handle multi-objective optimisation, for example, balancing maximum NPV with environmental metrics and integrating it directly into existing supervisory control systems, will be vital for actual field deployment.

Finally, the WOA-LSTM-SFS architecture demonstrates that combining artificial intelligence feature selection with advanced forecasting and optimisation algorithms is highly effective. It is computationally efficient, economically focused, and provides a highly practical blueprint for modernising gas injection strategies.

## REFERENCES

- Abolghasemi, S. S. S., Emadi, M., & Karimi, M. (2024). Accuracy improvement of breast tumor detection based on dimension reduction in the spatial and edge features and edge structure in the image. *Majlesi Journal of Electrical Engineering*, 18.
- Abdollahi, F., Hosseini, S., Sabet, M., Esmaceli-Faraj, S. H., & Amiri, F. (2021). A novel study of the gas lift process using an integrated production/injection system using artificial neural network approach. *Modeling Earth Systems and Environment*.
- Agwu, O. E., Alatefi, S., Azim, R. A., & Alkouh, A. (2024). Applications of Artificial Intelligence algorithms in Artificial Lift systems: a critical review. *Flow Measurement and Instrumentation*, 102613 .
- Alaidany, A. A., & Mahdi, M. M. A Review of IoT-Based Wearable Sensor Systems for Healthcare Monitoring .
- Ali A, A., Ali K, M., Marwah M, M., & Tibah, F. (2025). A Review of Machine Learning in Banking Risk Management and Possible Research TOPICS. *Journal of Engineering, Mechanics and Modern Architecture*, 4(1), 50–57 .



- Alsaedi, A. K. Z., Hussein, G. H., & Hamzah, F. N. (2025). Predicting Gas Oil Import Quantities in Iraq Using Machine Learning Techniques. *International Journal of Applied Mathematics*, 38(9s).
- Al-Shalabi, E. W., & Al-Hashem, M. (2021). Application of deep learning LSTM networks for reservoir performance prediction. *Fuel*, 306, 121702.
- Amir, A. H. A., & Nemer, Z. N. (2025). Inclusive Review on Advances in Masked Human Face Recognition Technologies. *arXiv preprint arXiv:2508.00841*.
- Amir, A. H. A., & Nemer, Z. N. (2024). Deep Learning-Based Siamese Neural Network for Masked Face Recognition.
- Dietrich, J., & Uniyal, B. (2020). Modeling interbasin groundwater flow in karst areas: Model development, application, and calibration strategy. *Environmental Modelling & Software*, 124, 104606 .
- Feng, Z., Huang, J., Jin, S., Wang, G., & Chen, Y. (2022). Artificial intelligence-based multi-objective optimisation for proton exchange membrane fuel cell: A literature review. *Journal of Power Sources*, 520, 230808 .
- Ghorbani, A., Khamehchi, E., & Mohammadi, A. H. (2022). Optimization of gas injection using PSO-based ANN models. *Energy*, 247, 123453.
- Huang, J., Li, Y.-F., & Xie, M. (2015). An empirical analysis of data preprocessing for machine learning-based software cost estimation. *Information and software Technology*, 67, 108-127 .
- Kalogirou, S. A. (2003). Artificial intelligence for the modeling and control of combustion processes: a review. *Progress in energy and combustion science*, 29(6), 515-566 .
- Karimi, M., Harouni, M., Jazi, E. I., Nasr, A., & Azizi, N. (2022). Improving monitoring and controlling parameters for Alzheimer's patients based on IoMT *Prognostic models in healthcare: Ai and statistical approaches* (pp. 213-237): Springer
- Karimi, M., Karimi, Z., Khosravi, M., Delaram, Z., Dehsheikhim, M. H., Najafabadi, S. A., . . . Tavakoli, N. (2025). Feature Selection Methods in Big Medical Databases: A Comprehensive Survey. *International Journal of Theoretical & Applied Computational Intelligence*, 181-209 .
- Katterbauer, K., Al Qasim, A., Al Shehri, A., & Yousif, A. (2022). *A novel artificial intelligence framework for the optimal control of wireless temperature sensors for optimizing oxygen injection in subsurface reservoirs*. Paper presented at the Offshore Technology Conference Asia.
- Kaur, S., Kumar, Y., Koul, A., & Kumar Kamboj, S. (2023). A systematic review on metaheuristic optimization techniques for feature selections in disease diagnosis: open issues and challenges. *Archives of Computational Methods in Engineering*, 30(3), 1863-1895 .
- Krishnan, A., Sundaram, T., Nagappan, B., & Devarajan, Y. (2024). Integrating artificial intelligence in nanomembrane systems for advanced water desalination. *Results in Engineering*, 103321 .



- Li, Z., Sun, Q., & Mohanty, K. K. (2023). *Multivariable prediction of reservoir behavior using recurrent neural networks*. *Journal of Petroleum Science and Engineering*, 220, 111292.
- Mahdi, M. A. A. (2023). *Integrated data analysis approach to select artificial lift method using machine learning*.
- Malozymov, B. V., Martyushev, N. V., Kukartsev, V. V., Tynchenko, V. S., Bukhtoyarov, V. V., Wu, X., ... & Kukartsev, V. A. (2023). Overview of methods for enhanced oil recovery from conventional and unconventional reservoirs. *Energies*, 16(13), 4907.
- Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software*, 95, 67-51.
- Moelyono, S., Misra, S., Retnanto, A., Palviriyachote, S., Voneiff, G., & Thompson, A. (2025). *Utilizing Machine Learning for Gas Lift Optimization: Reducing Interventions and Accelerating Production*. Paper presented at the International Petroleum Technology Conference.
- Movahed, P., Burmaster, D., Karantinos, E., Villarreal, A., Memarzadeh, M., Vela, S., . . . Banes, T. (2024). *Gas Lift Optimization In The Permian Using Machine Learning And Artificial Intelligence*. Paper presented at the SPE Artificial Lift Conference and Exhibition-Americas?
- Olszak, E., Yoho, M., Greaser, R., Srivastava, R., Lanke, R., Moffett, R., . . . Hingerl, F. (2022). *Case Study: Plunger Lift Optimization Leveraging Physics-Assisted AI and the Impact on Greenhouse Gas Emissions Reductions*. Paper presented at the SPE Annual Technical Conference and Exhibition?
- Ramos, M., Khalifah, M., Almilad, R., & Mishkhes, A. (2024). *Optimization of Oil Production Through Advanced Gas Lift Technology*. Paper presented at the SPE Artificial Lift Conference and Exhibition-Americas?
- Rana, N., Latiff, M. S. A., Abdulhamid, S. i. M., & Chiroma, H. (2020). Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments. *Neural Computing and Applications*, 32, 16245-16277.
- Shokir, E. M. E.-M., Hamed, M. M., Ibrahim, A. E.-S., & Mahgoub, I. (2017). Gas lift optimization using artificial neural network and integrated production modeling. *Energy & Fuels*, 31(9), 9302-9307.
- Temizel, C., Canbaz, C. H., Betancourt, D., Ozesen, A., Acar, C., Krishna, S., & Saputelli, L. (2020). *A comprehensive review and optimization of artificial lift methods in unconventional*. Paper presented at the SPE Annual Technical Conference and Exhibition?
- Wang, L., Yao, Y., Luo, X., Adenutsi, C. D., Zhao, G., & Lai, F. (2023). A critical review on intelligent optimization algorithms and surrogate models for conventional and unconventional reservoir production optimization. *Fuel*, 350, 128826.



- Waqar, A., Othman, I., Shafiq, N & ,.Mansoor, M. S. (2023). Applications of AI in oil and gas projects towards sustainable development: a systematic literature review. *Artificial Intelligence Review*, 56(11), 12771-12798 .
- Zeinolabedini, L., Ameli, F., & Hemmati-Sarapardeh, A. (2025). Integration of Artificial Intelligence and Advanced Optimization Techniques for Continuous Gas Lift under Restricted Gas Supply: A Case Study. *Digital Chemical Engineering*, 100220 .
- Zhang, Y., Erfani, R., & Elkamel, A. (2020). *A machine learning framework for optimizing oil production in petroleum reservoirs*. *Journal of Petroleum Science and Engineering*, 195, 107732.
- Zhou, Y., Kang, J., Kwong, S., Wang, X., & Zhang, Q. (2021). An evolutionary multi-objective optimization framework of discretization-based feature selection for classification. *Swarm and Evolutionary Computation*, 60, 100770 .
- Zhu, D., & Furui, K. (2006). *Optimizing oil and gas production by intelligent technology*. Paper presented at the SPE Annual Technical Conference and Exhibition?