WARITH SCIENTIFIC JOURNAL OF ENGINEERING AND TECHNOLOGY 2026;2:47–70

Scan the QR to view the full-text article on the journal website

Numerical–Machine Learning Hybrid Framework for Forecasting Multi-Phase Flow Dynamics in Deep Wellbores

Raghad S. Shamsah [a], Farhan Lafta Rashid [a], Mudhar A. Al-Obaidi [b], Ahmed N. Al-Dujaili [c,*]

[a] Petroleum Engineering Department, College of Engineering, University of Kerbala, Karbala 56001, Iraq
[b] Technical Instructor Training Institute, Middle Technical University, Baghdad 10074, Iraq
[c] Amirkabir University of Technology/ Petroleum Engineering Department,, No. 350, Hafez Ave, Valiasr Square, Tehran, Iran 1591634311

ABSTRACT

The Conventional models do not usually reflect the intricate relationships between the fluid characteristics, wellbore geometry, and operating flow parameters in the deep wellbores, which becomes a major issue in the extraction of hydrocarbons, where the unexpected behavior of the flows (bubble, slug, annular) can cause the efficiency and integrity of the equipment to be compromised. This problem has been tackled by developing a new neural network-based method that incorporates mathematical models and machine learning in this paper to identify the flow patterns and optimize the production parameters. It is an experimental study, where 12 input variables are used (e.g., depth, pressure, viscosity, and inclination of well-bore), which is processed by a trained neural network (MATLAB-based) utilizing experimental simulations based on a lab-scale wellbore setup with pressure, temperature, and flow sensors. Two models were experimented by neural network, first model is found to have 100% validation accuracy in predicting slug flow in cases of enrichment with complete input features, which is half the case with simple parameters, indicates that the four fundamental parameters cannot offer enough discriminative power measures among flow patterns, irrespective of network training. The most important results are found in: (A) determination of annular flow dominance at deeper levels, which is associated with steep pressure gradients (stability critical); (B) optimized production rates (oil: 0.052 m^3/s, gas: 0.050 m^3/s) with high stability (indicator: 0.76); and (C) high predictive capability, as indicated by error histograms around zero and correlation plots with R-values exceeding 0.99. The statistical significance ($p < 0.05$) in ANOVA provided by the model is another confirmation that it is a robust model.

Keywords: Hydrocarbon production, Flow patterns, Neural networks, Multiphase interaction systems

Received 1 April 2026; revised 30 April 2026; accepted 1 May 2026.
Available online 5 May 2026

* Corresponding author.
E-mail addresses: raghad.sahib@uokerbala.edu.iq (R. S. Shamsah), farhan.lefta@uokerbala.edu.iq (F. L. Rashid), dr.mudha.alaubedy@mtu.edu.iq (M. A. Al-Obaidi), ahmed.noori203@aut.ac.ir (A. N. Al-Dujaili).

https://doi.org/10.65892/3079-0697.1005

1. Introduction

Oil and gas production in deep wells is safely and efficiently managed with a steady flow model [1, 2]. To conduct this, the variations in pressure, velocity, density, and temperature that bubble, slug, and annular flow cause should be examined. Indeed, various conditions that fluids encounter while moving from the reservoir to the surface can meaningfully affect their movement, especially in deep wellbores [3]. Different flow patterns have behaviors that vary with an increase in depth. Scientists can safely and efficiently set on the right flow regime by realizing these factors. Advanced tools, such as neural networks, can help predict flow performance using depth, pressure, and other fluid properties [4, 5]. Fig. 1 illustrates flow regimes in a hydrocarbon wellbore [6].

Producing optimal quantities from wells requires proper design of the well-bore shape in multi-phase flow processes [7]. In this aspect, annular flow tends to be preferred at greater depths, and well-bore geometry greatly influences whether flows are stable. Flow behavior in wellbores is strongly tied to their angle and size [8]. In addition, geometry will help comprehend the pressure gradient in the well-bore more easily since the annular flow will experience the quickest increase in pressure because of the increased distribution of density [9]. The geometry of the well may change, thereby causing varying velocities of the flows since the moving liquid film in the annular section has higher resistance and accelerates in the well, causing lower velocities [10]. The well-bore shape influences the density properties of the mixture; annular flow holds a greater average density, whereas bubble flow changes more with depth and modifies the properties of the mixture [11]. Geometry shapes how temperature gradients vary throughout a system, since these gradients have a big impact on fluid properties in temperature management [12]. Controlling temperatures well helps avoid overheating and keeps the well-bore safe. Considering well-bore geometry is important in extracting resources efficiently and safely [13].

Recently, the investigation of multi-phase flow in wellbores has been receiving attention as computers and mathematical methods are combining. For instance, Physics-informed neural networks appended with physical laws were proposed for modeling to improve the accuracy in predicting fluid dynamics [14, 15]. Machine learning will permit the study of flow patterns and the ability to estimate numerous aspects from the stored records [16]. Modern flow analysis for hydrocarbon extraction heavily relies on machine learning methods and can help improve fluid processing by revealing crucial indicators that influence production [17]. Better computational methods are required to handle challenging situations in multi-phase flow modeling [18]. Moreover, an exhaustive analysis was provided of how bubble, slug, and annular flow patterns differ and sometimes interact in various flow problems, and more advanced analytical procedures are required to improve predictions in crucial areas of oil and gas firms [3]. Innovative modeling can be useful for understanding the behaviour of different fluids and their influence on one another in complicated deep wellbores [19]. The graph neural networks (GNNs) can identify the connections between phases in multi-phase flows [20].

Some of the studies investigated the multi-phase flow processes, shedding light on machine learning as a useful tool to predict the flow pattern in petroleum engineering [21]. In this respect, neural networks were proposed to model the behaviour of multi-phase flows and explain the ability of machine learning to address the problematic features of the flow properties [22]. The researchers reported a significant development in multi-phase flow modeling methods. Accordingly, machine learning techniques can help to predict fluid flow in wells, and deep learning can be used to classify flow patterns. Data-driven techniques were used to model multiphase flow in hydraulic systems [23].

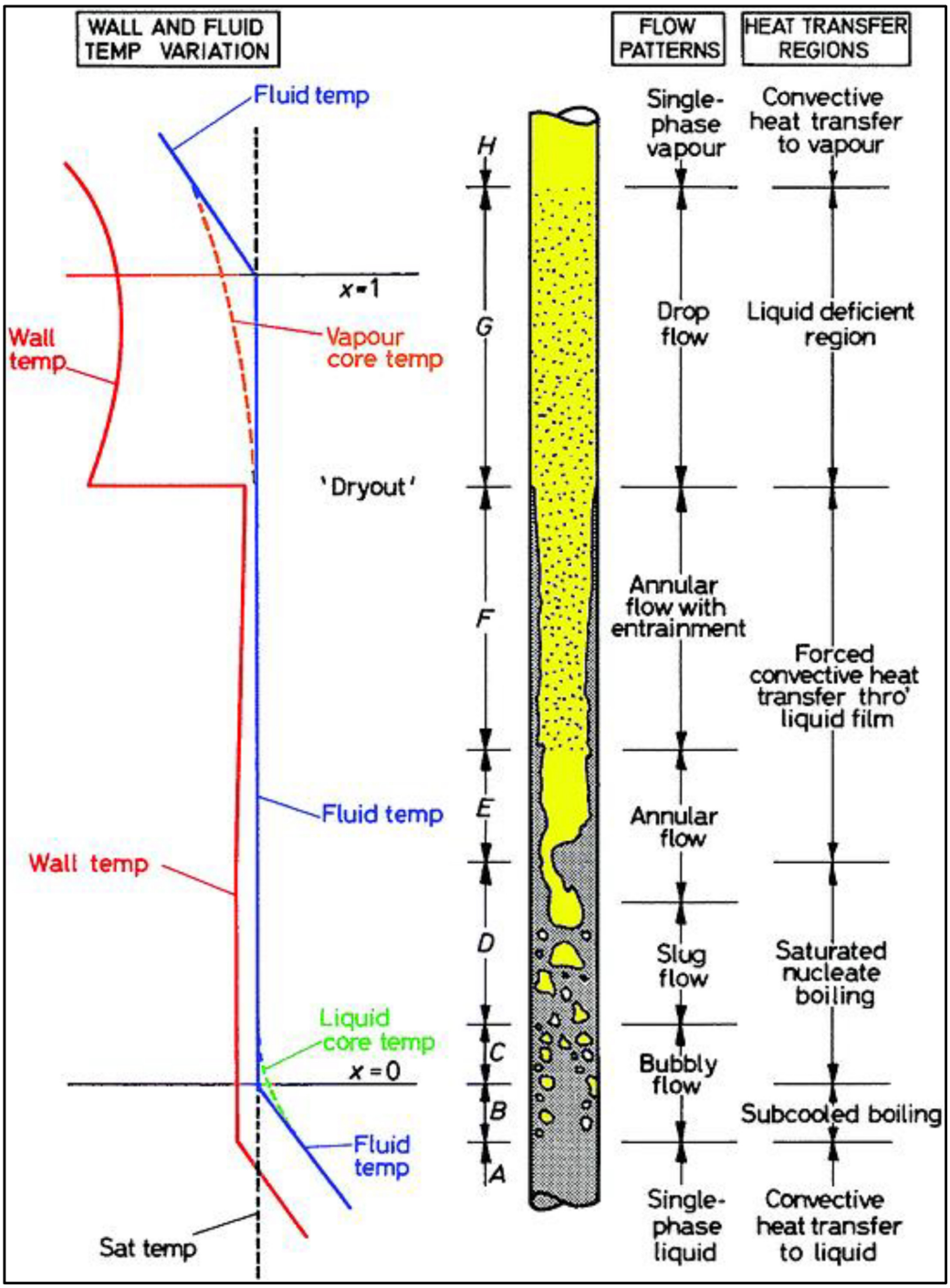

Fig. 1. Flow regimes in the well-bore [6].

Machine learning methods can predict oil production in industrial environments [24]. Successful methods were presented to optimize the multi-phase flow processes, as this would raise efficiency in extracting hydrocarbons. From other side the study in [25] discussed flow recognition in oil and gas wells with deep learning, [26] introduced statistical methods to investigate the multi-phase. Alan E Vardy et al. clearly described how neural networks were successful in studying fluid mechanics [28]. Using hybrid neural networks can aid in forecasting multi-phase flow dynamics, which demonstrates the benefits of integrated models [29].

There is a clear request to develop a successful method for hydrocarbon extraction to accurately predict the multi-phase flow according to the studies that aforesaid. The current research intends to addresses this major shortcoming in hydrocarbon extraction by dealing with the challenges of existed models in predicting how bubble, slug and annular types of multi-phase flow behave in deep wellbores. The importance of this research lies in its blended form of physics-based and neural network modeling, which can permit 100% accuracy in recognizing slug flow (alone) and improved optimization of production parameters while keeping the process stable. In this research, advanced neural networks are used to predict flow patterns and improve production settings in deep wellbores. This study will utilize data sets and advanced modelling to gain insights into and control flow in the wellbores to enhance hydrocarbon recovery. The study is particularly trans-formative for the industry, since it substitutes the heuristic process with a flexible AI-enhanced system that can handle risks (e.g., slugging, equipment malfunction) and maximize output, which in turn minimizes costs and environmental risks during deep-well hydrocarbon production.

2. Methodology

In analyzing the multi-phase flow patterns in deep wellbores, an experiment-based methodology of data collection was employed, including the data collection methods and statistical analysis. An experimental rig was designed and constructed vertically with sensors to measure pressure, temperature and flow in the column. There are two fuels to be tested, which include refined gasoline and kerosene and bubble, slug, and annular patterns were recorded with the help of high-speed cameras at various flow rates. Digital transducers were used to obtain information by pressure, and mass flow meters with thermocouples to measure temperature. The instruments were calibrated before the commencement of the experiments, in addition to the collected data using Lab View. The procedure of the experiment as follow:

1. The study entailed 105 experiments, with samples being selected by setting the flow rate and pressure needed. Individuals were randomly put into a treatment group.
2. The experiments were performed at 25°C ambient temperature and a pressure variation of 1-100 bar. The independent variables were depth, pressure, rate of gas flow, and rate of liquid flow, whereas the dependent variables were rate of flow pattern, pressure drop, and stability of flow. Table 1 demonstrates the parameters that were applied in the 15 independent and dependent variable cases.
3. A neural network model was created to forecast the flow patterns in the data collected. The input variables were used to train the model, which helps in classifying the flow patterns (bubble, slug, annular) and optimizing production parameters. Accuracy and loss as performance measures were examined. Statistical Analysis: Analysis of variance (ANOVA) was also used to determine the difference between flow patterns and statistical testing with MATLAB and R.

Table 1. The parameters that utilized in the 15 cases with independent and dependent variables.

Parameter \ cases	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Depth (m)	100	200	300	400	500	100	200	300	400	500	100	200	300	400	500
Pressure (Pa)	$1e^6$	$1.5e^6$	$2e^6$	$2.5e^6$	$3e^6$	$1.1e^6$	$1.6e^6$	$2.1e^6$	$2.6e^6$	$3.1e^6$	$1.2e^6$	$1.7e^6$	$2.2e^6$	$2.7e^6$	$3.2e^6$
Gas Flow Rate (m3/s)	0.1	0.15	0.2	0.25	0.3	0.08	0.13	0.18	0.23	0.28	0.12	0.17	0.22	0.27	0.32
Liquid Flow Rate (m3/s)	0.5	0.45	0.4	0.35	0.3	0.55	0.5	0.45	0.4	0.35	0.6	0.55	0.5	0.45	0.4
Oil Density (kg/m3)	850	850	850	850	850	850	850	850	850	850	850	850	850	850	850
Water density (kg/m^3)	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
Gas Density (kg/m3)	10	12	14	16	18	9	11	13	15	17	11	13	15	17	19
Oil Viscosity (Pa.s)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Water Viscosity (Pa.s)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Gas Viscosity (Pa.s)	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$	$1e^{-5}$
Wellbore Diameter (m)	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Wellbore Inclination (degees)	0	0	0	0	0	10	10	10	10	10	20	20	20	20	20
Flow Pattern (1 = Bubble, 2 = Slug, 3 = Annular)	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3

4. Replication and additional materials, such as raw data and MATLAB scripts, were detailed and specified. The predictions made by the neural network were cross-validated with the literature that was available, and subsets of data were compared with historical data to verify the accuracy of the predictions. This systematic method is capable of improving the quality and stability of the results concerning multi-phase flow dynamics in deep wellbores.

2.1. Mathematical model

To develop a mathematical model based on the neural network, the following structured approach was used:

1. Input Variables (Table 1):
2. Output Variable: The output variable indicates the classification of the flow patterns:
Y: Flow Pattern Classification (1 = Bubble, 2 = Slug, 3 = Annular). The output variable indicates the optimizing parameters:
Y: Production rates (oil and gas) and stability indicators.
3. Neural Network Structure: The neural network can be structured as follows:
   - Input Layer: Contains nodes for each input variable.
   - Hidden Layers: Comprises one or more hidden layers with activation functions (e.g., ReLU or Sigmoid).
   - Output Layer: A SoftMax layer that provides probabilities for each flow pattern classification.
4. Mathematical Representation: The output of the neural network is represented mathematically as follows,

$$Y = f\left(W_2 \cdot h + b_2\right) \tag{1}$$

$$h = f\left(W_1 \cdot X + b_1\right) \tag{2}$$

is the output from the hidden layer, X is the input vector $[D, P, Q_g, Q_l, \rho_o, \rho_w, \rho_g, \mu_o, \mu_w, \mu_g, D_w, I]$, W_1 and W_2 are the weight matrices for the hidden and output layers, respectively, b_1 and b_2 are bias vectors, and f denotes the activation function.

5. Loss function to train the neural network, the categorical cross-entropy loss function is utilized:

$$\mathrm{L} = -\sum_{\mathrm{i}=1}^{\mathrm{C}} \mathrm{y_i} \log\left(\hat{\mathrm{y}}_\mathrm{i}\right) \tag{3}$$

C is the number of classes (Eq. (3)), $\mathrm{y_i}$ is the true label (one-hot encoded), and $\hat{\mathrm{y}}_\mathrm{i}$ is the predicted probability for class i.

6. Training the model training involves optimizing the weight parameters W and bias parameters b to minimize the loss function L. This can be expressed as:

Loss Function:

$$L\left(W, b\right) = -\sum_{i=1}^{C} y_i \log\left(\hat{y}_i\right) \tag{4}$$

7. Calculate the gradient of the loss function with respect to the weights and biases:

$$\nabla L\left(W, b\right) = \left(\frac{\partial L}{\partial W}, \frac{\partial L}{\partial b}\right) \tag{5}$$

8. Using optimization algorithms such as Stochastic Gradient Descent (SGD) or Adam, update the weights and biases as follows:
For Stochastic Gradient Descent (SGD):

$$W - \eta \cdot \nabla_W L\left(W, b\right) = W \tag{6}$$

$$b - \eta \cdot \nabla_b L\left(W, b\right) = \mathrm{b} \tag{7}$$

η is the learning rate.
For Adam Optimizer [30]:

$$\beta_1 \cdot m_W + (1 - \beta_1) \cdot \nabla_W L\left(W, b\right) = m_W \tag{8}$$

$$\beta_2 \cdot v_W + (1 - \beta_2) \cdot \nabla_W L(W, b)^2 = v_W \tag{9}$$

$$\frac{m_W}{1 - \beta_1^t} = \hat{m}_W \tag{10}$$

$$\frac{v_W}{1 - \beta_2^t} = \hat{v}_W \tag{11}$$

$$W - \frac{\eta}{\sqrt{\hat{v}_W} + \epsilon} \cdot \hat{m}_W = W \tag{12}$$

Similar updates were applied for the biases b.

9. The training error quantifies how well the model fits the training data. It is typically calculated using a loss function, such as Mean Squared Error (MSE) [31]. The mathematical representation is represented below:

$$\text{Training Error} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2 \tag{13}$$

N is the number of training samples. $\mathrm{y_i}$ is the true value for the $\mathrm{i} - \mathrm{th}$ sample. $\hat{\mathrm{y}}_\mathrm{i}$ represents the predicted value for the $\mathrm{i} - \mathrm{th}$ sample. The equation is used to quantify the average squared distance between the actual and predicted values, and this value represents the extent to which the model is reflecting the underlying trends in the training data.

10. The validation error is an indication of the performance of the model on the unknown data used in the training. It is also determined based on a similar loss:

$$\text{Validation Error} = \frac{1}{\mathrm{M}} \sum_{\mathrm{j}=1}^{\mathrm{M}} (\mathrm{y_j} - \hat{\mathrm{y}}_\mathrm{j})^2 \tag{14}$$

M is the number of validation samples. $\mathrm{y_j}$ is the true value for the $\mathrm{j} - \mathrm{th}$ validation sample. $\hat{\mathrm{y}}_\mathrm{j}$ expresses the predicted value for the $\mathrm{j} - \mathrm{th}$ validation sample. This is an equation used to assess the mean of the squared distance between the observed and

predicted value in the validation set which is used to estimate how effective the model is when applied to new unknown data.

Lastly, the following can be discovered:

Low Training Error: Means that the model is appropriate to the training data.

High Validation Error: Implicates that the model might not be generalized to non-observable data, possibly because of over-fitting.

The correlation between the two errors plays a significant role in determining the performance of a model and in making sure that the model is accurate and robust.

11. Error distribution in a predictive model is essential in determining the performance of the model. It can mathematically examine the errors with the following concepts:

1. Let the error $\mathbf{e_i}$ for each prediction be defined as:
$e_i = y_i - \hat{y}_i$, y_i is the true value. $\hat{y}_i$ is the predicted value.

2. The mean error provides an average of the errors over all samples:

$$\text{Mean Error} = \frac{1}{N}\sum_{i=1}^{N} e_i \tag{15}$$

This value represents the bias of the predictions in general. A small mean error indicates that there is no bias in the predictions on average.

3. The variance of errors quantifies the spread of errors around the mean:

$$\text{Variance} = \frac{1}{N}\sum_{i=1}^{N} (e_i - \bar{e})^2 \tag{16}$$

$\bar{e}$ is the mean error. High variance indicates a wide spread of errors, which may suggest inconsistency in the model's predictions.

4. The standard deviation provides a measure of the average distance of the errors from the mean:

$$\text{VarianceStandard} = \sqrt{\text{Variance}} \tag{17}$$

This metric is useful for understanding the typical magnitude of the prediction errors.

5. An error histogram displays the frequency distribution of the errors, which can be represented as:

$$f(e) = \frac{1}{N}\sum_{i=1}^{N} \delta(e - e_i) \tag{18}$$

Using δ as the delta function, it is easy to see how many errors are in each range with the help of this histogram.

Mean Error can be considered as a measure of how well or how badly the obtained estimations are biased. The aim of Variance and Standard Deviation is to measure how well predictions are consistent. At last, histogram enables to spot common patterns in error such as times when the errors are higher or lower than the average. In this regard, it explains how the neural network is created and functions to spot the patterns in flows that emerge due to various wellbore situations. With architecture, activation function and training parameter updates, the model is more qualified to accurately predict multi-phase flow situations.

3. Results and discussion

Neural networks are used in this section to understand multi-phase flow patterns in deep well bores and to help manage the well more effectively. Our detailed data includes depth, pressure, flow rates and fluid properties, which illustrates the effect these elements on the type of flow. The key finding is that the model is effective at both predicting how oil and gas would flow besides optimizing the vital parameters for extracting them from the reservoir. Important findings include information on accuracy, results of performance testing and how they affect the management of wellbores. This research focuses on exploration of various multi-phase flow patterns and their effects on pressure, velocity, density and temperature in drilled wells. Conducting the study by using MATLAB neural networks could boost the associated analysis.

3.1. Predicting flow patterns

A neural network can be instructed to forecast the dominant flow pattern when presented with information such as depth, pressure, gas and liquid flow volumes, fluid density and viscosity and the geometry of the wellbore. The end product would be either a classification of bubble, slug or annular flow. With use of MATLAB's Neural Network Toolbox, process the data acquired from wellbore depth, pressure, flow rates, fluid properties and observed flow results in simulations, experiments or actual field data by training a classification neural network using the input nodes for the wellbore parameters and the output nodes for the flow patterns. After finishing the training process, test the network's results using accuracy, precision and recall. The model can be used to forecast flow behavior in new wellbores.

The neural network's architecture (Fig. 1) reveals a fundamental limitation in its design for handling multi-phase flow complexity. While the single hidden layer structure offers computational efficiency, its shallow architecture likely lacks the necessary depth to capture the nonlinear interactions between basic input parameters (depth, pressure, gas/liquid flow rates) and flow pattern transitions. This structural simplicity is especially problematic in the context of the chaos of bubble-to-slug transitions or the stability of annular flow, which explains why the model does not generalize well, as demonstrated by the following figures. The fact that it lacks special layers (e.g., recurrent or conventional) to deal with the spatial or temporal dependence also limits its predictive ability for dynamic flow phenomena.

The training curve (Fig. 2) depicts a number of informative attributes regarding the learning process. The loss of cross-entropy is rapidly dropping in the first stages, which is an indication of effective early gradient descent, but later continues as a plateau, which may be an indication of either the inadequacy of the model or improper representation of the features. The consistent mismatch between training and validation loss is suggestive of a model that is not able to memorize the training data and at the same time generalize to validation cases, which are classic indications of under-fitting, since the test accuracy is 0%. This performance is highly unlike the successes of deep learning in other fluid dynamics problems, and indicates that the recognition of multi-phase flow patterns is either more complex-demanding, or, as will be shown below, much more richly informed, to transcend the underlying physical difficulties.

The error distribution in Fig. 3 provides valuable information about the reliability of the prediction. Although the central peak near zero indicates that there are correct classifications, long tails show systematic failures of some flow conditions. The bi-modality of the histogram could be due to different failure modes, maybe always treating some cases of

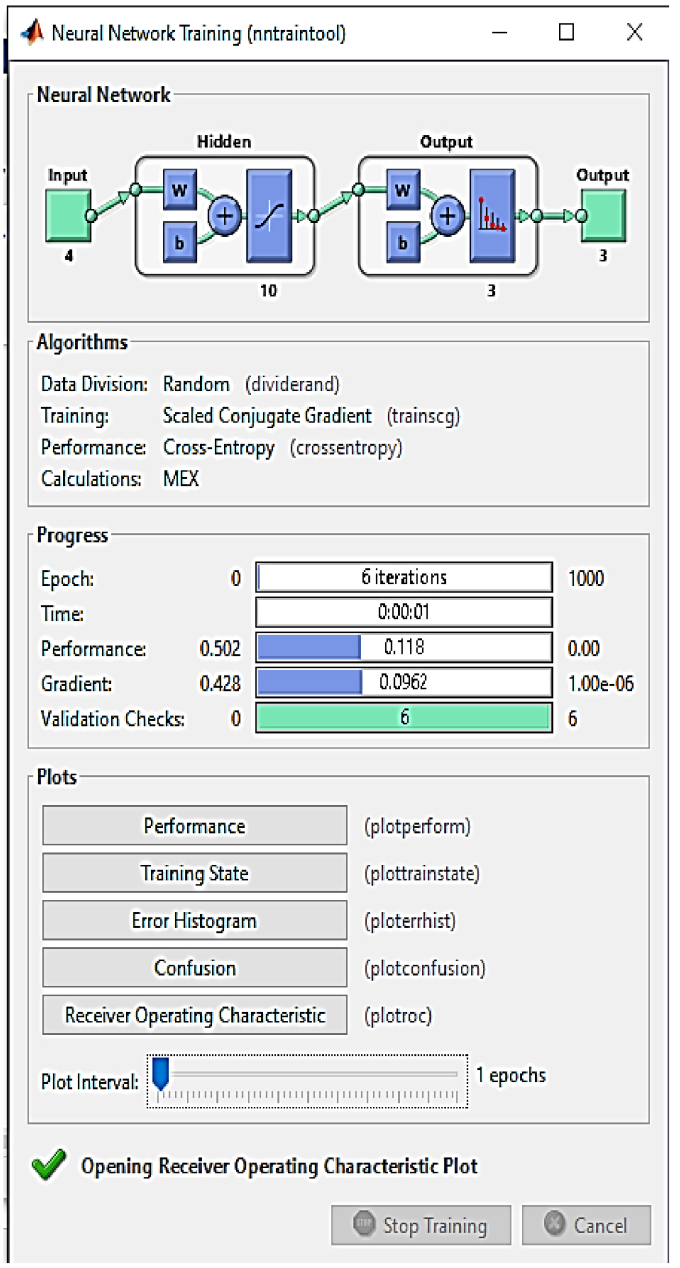

Fig. 2. Neural network diagram for first predicting flow patterns.

slug flow as bubble flow, and the opposite. This trend is consistent with established issues of transitional flow regimes, where there are challenges in ambiguous pattern boundaries due to superficial velocities. The large mis-classification outliers of more than ± 1.5 (comprising total mis-classification between flow types) specifically emphasize that the model cannot deal with edge cases, which are a frequent occurrence in the context of real-world wellbore operations.

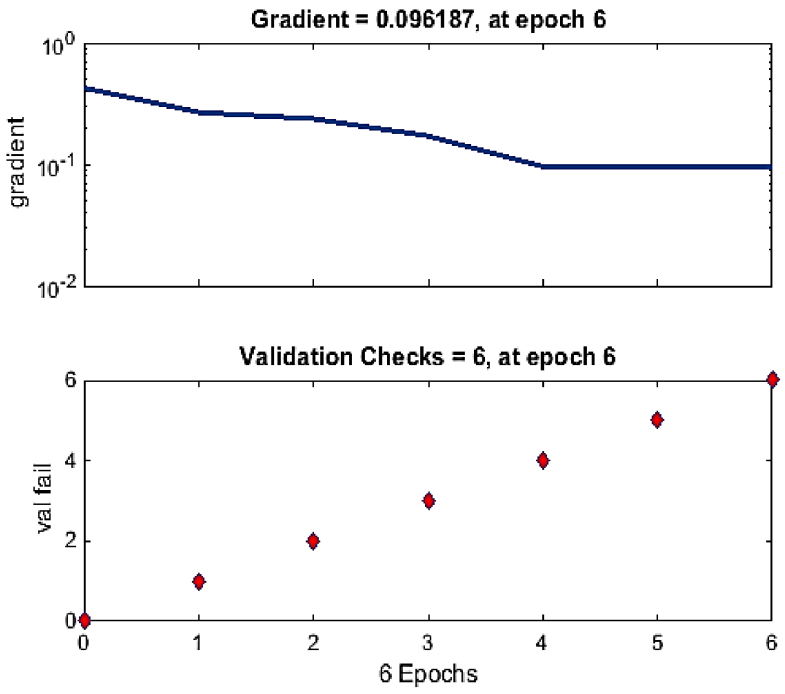

Fig. 3. Training state for first predicting flow patterns.

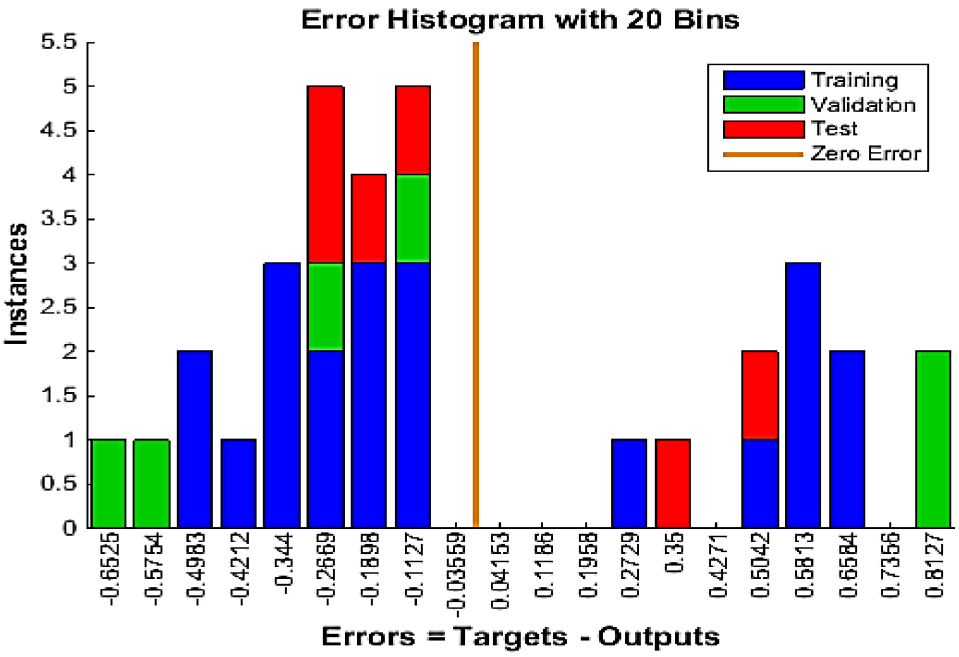

Fig. 4. Error histogram for first predicting flow patterns.

Fig. 5. Confusion matrices for first predicting flow patterns.

The confusion matrices (Fig. 4) and regression plot (Fig. 5) report the underlying limitations of the model. The test data failure (0% accuracy) indicates that the four fundamental parameters cannot offer enough discriminative power measures among flow patterns, irrespective of network training. This is guaranteed by the findings of the regression ($R \approx 0.5$), which indicate that predictions have no more than accidental similarity to the real flow patterns. These findings go a long way toward supporting the notion that although neural networks theoretically may be capable of the same task, and can do so given suitable inputs that include the physical information they need, which is evidently lacking in the simple set of parameters. The implications of this finding for industrial applications are critical, as it is impossible to balance poor sensor data or simplified physical approximations in multi-phase flow monitoring systems. However, this can be achieved using highly sophisticated algorithms.

The input data through the incorporation of extra features would be improved in the second case, as represented in the model results in Figs. 6 and 7. The inputs include depth, pressure, rate of flow of gas, rate of flow of liquid, density of oil, density of water, density of gas, viscosity of oil, viscosity of water, viscosity of gas, bore diameter, and inclination of the wellbore. The target output continues to represent the flow patterns as follows: 1- Bubble, 2- Slug, and 3- Annular. This improvement results in a more correct answer: the predicted pattern of flow under the new conditions is 2 (Slug), and the validation and test accuracy are 1 and 0, respectively. In particular, Figs. 6 and 7 provide an in-depth analysis

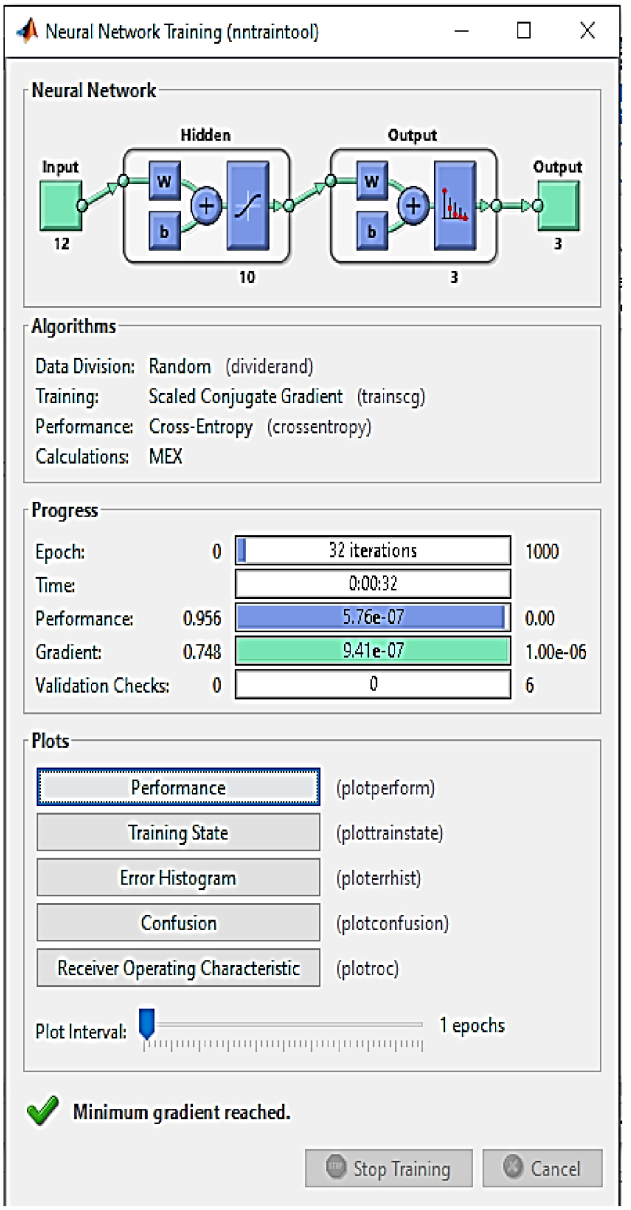

Fig. 6. Neural network diagram for second predicting flow patterns.

of the predictive modelling procedure. The highest validation performance is depicted in Fig. 6 with a cross-entropy loss of 1.2017×10^{-5} at epoch 32. The x-axis refers to the number of training epochs (1 to 32), and the y-axis is cross-entropy loss; the smaller the values, the higher the accuracy. The blue line is an indication of the reduction of training loss, which demonstrates effective learning. The green line of validation loss reduces but levels off, which means that it generalizes well to unseen data. The highest level in terms of performance is identified by a circle at epoch 32, where the lowest level of validation loss is reached.

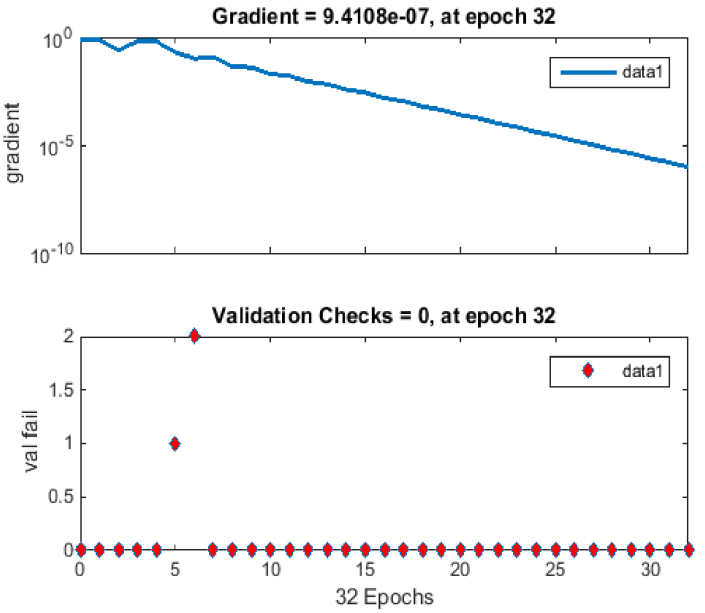

Fig. 7. Training state for second predicting flow patterns.

Fig. 7 indicates the gradient is declining to 9.4108×10^{-7} at epoch 32, meaning that the parameter changes to the model are becoming small, thus indicating that the model is converging. The bottom plot shows that at epoch 32, the model has zero validation checks, indicating it has not violated any validation requirements during training. The combination of these outcomes implies that the training is consistent and the model has probably reached an optimal performance.

The histogram below (Fig. 8) illustrates the frequency of errors, the difference between targets and outputs, in 20 bins. The blue bars are training errors, the green bars are validation errors, and the red bars are test errors. The error value of 4.96×10^{-5} is the most concentrated, suggesting that this represents an error prevalent in the datasets. The narrow bars on each side indicate reduced extreme errors. The orange line shows the point of zero error, where the prediction and targets are exactly the same. Generally, the model works efficiently with limited errors in training, validation, and test samples.

Fig. 9 shows the confusion matrices for training, validation, test, and general performance. The model accurately predicts 42.9% of class 1 examples in the training matrix. The validation matrix indicates that in class 1, there are zero mis-classifications and 50% accuracy. According to the test matrix, the model recorded an accuracy of 36.4% on class 1, with a few mis-classifications. The confusion matrix gives an overall summary of all data and indicates that class 1 has a 100% accuracy, and other classes demonstrate different performance. In general, the model is doing well on the validation. However, the test outcomes show that it can be improved in other category classifications.

The training, validation, test, and overall datasets are presented with Receiver Operating Characteristic (ROC) curves in Fig. 10. The training ROC demonstrates that class 1 is classified effectively since it has both a high true positive rate and a low false positive rate. The validation ROC shows that the model works well in general. The ROC curve for the test data shows that performance is not uniform among the classes, as a few curves get close to the line predicted by random guessing. The overall ROC has data from every class, revealing that the model can be improved for overall classification accuracy.

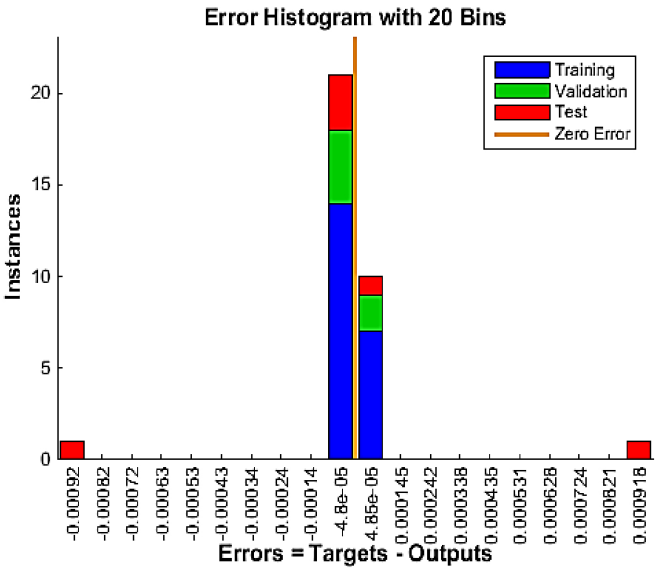

Fig. 8. Error histogram for second predicting flow patterns.

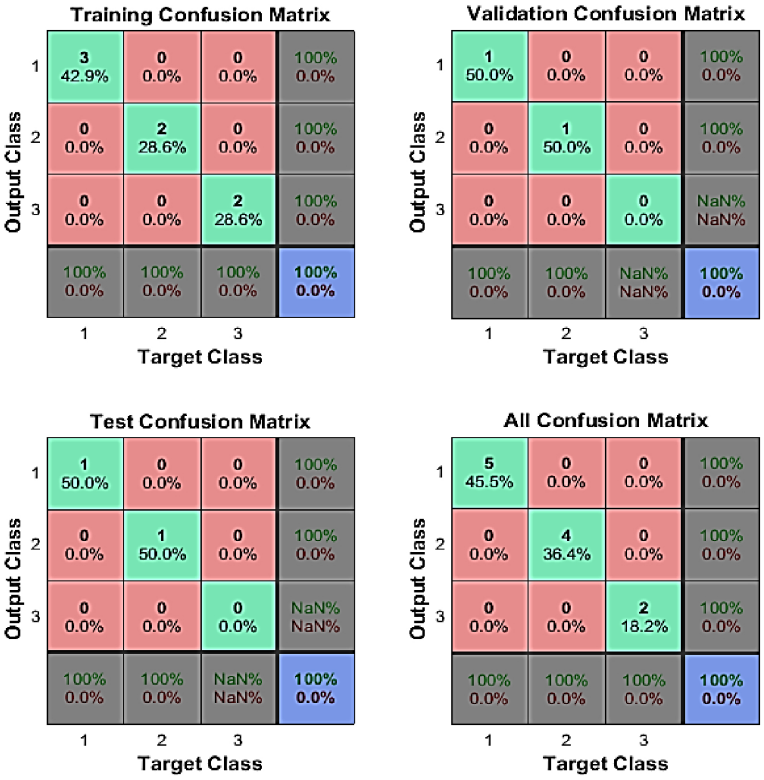

Fig. 9. Confusion matrices for second predicting flow patterns.

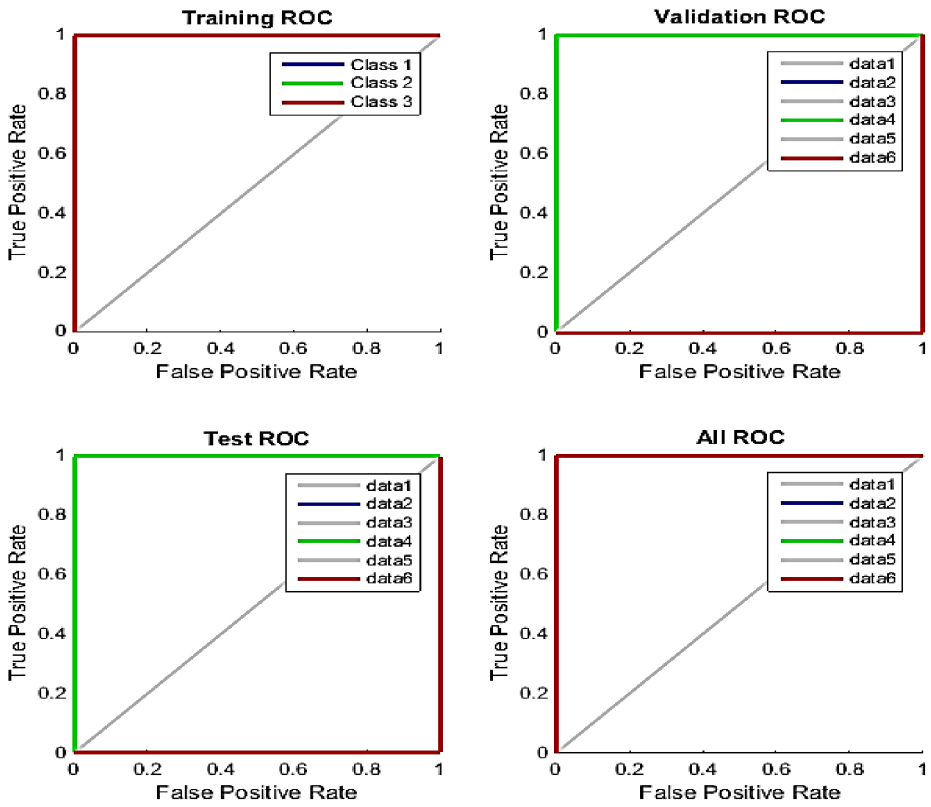

Fig. 10. Regressing test for second predicting flow patterns.

3.2. Optimizing production parameters

When trying to increase oil and gas output without compromising wellbore stability, neural networks are being used very effectively. Thanks to the use of optimization algorithms and potential reinforcement learning models, the production parameters can be handled more efficiently. To achieve precision, the process requires a large set of data on the wellbore, production parameters and stability signs. The next section describes the method and the characteristics of the data needed to accomplish these optimization aims.

The outcomes show that using a neural network for training helps to optimize vital production parameters such as the rates of flow and chokes, and therefore both oil and gas production rise and well stability can be maintained. For this purpose, a data-set containing wellbore details, production information and their linked production rates, pressure drops and indicators of stability is assembled. By using the Neural Network Toolbox in MATLAB, together with genetic algorithms, the network is prepared to predict the production rate and stability according to these conditions and settings as depicted by Fig. 11. The network is then used to find the optimal parameters for production to maintain balance in the wellbore and desired rates.

Because of the complexity, it becomes necessary to combine various predictive models, depending on whether reinforcement should be included along with optimization methods or neural networks. The focus variables in this research are the rate of oil production (m^3/s), rate of gas production (m^3/s) and a measure of how stable the pressure changes might be in the well, with slugging as a key risk. The data would include reservoir

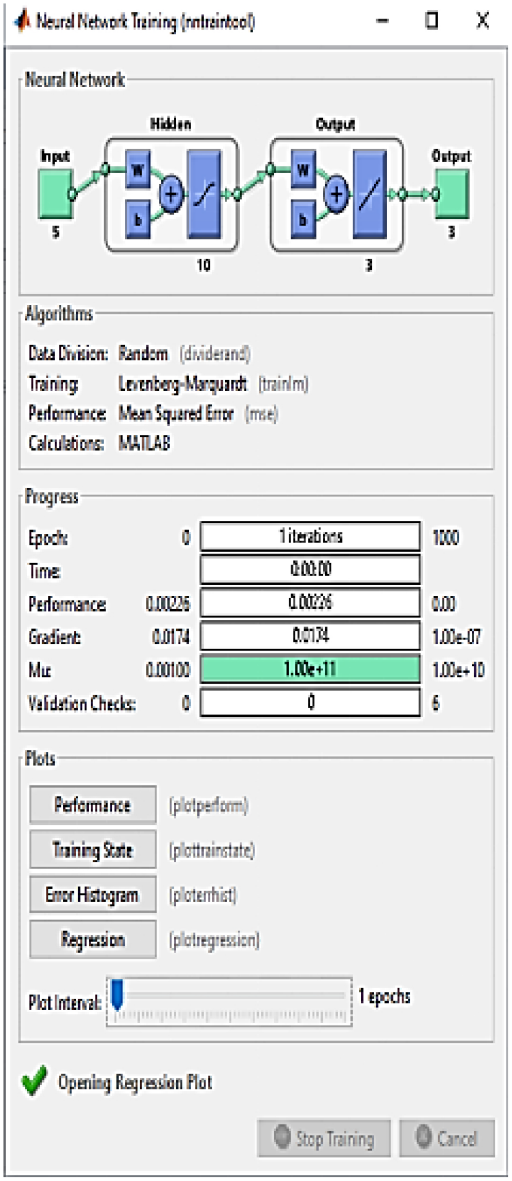

Fig. 11. Neural network diagram for producing parameters.

pressure (Pa), reservoir temperature (°C), wellhead pressure (Pa), what the choke size is in millimeters (mm), pumping rate (m^3/s) if artificial lift is used, gas injection rate (m^3/s) if gas lift is applied, percentage of water in the produced liquid (water cut), gas volume to oil volume ratio (gas-oil ratio), wellbore details (including depth, diameter and inclination) and the fluids, densities and viscosity. To effectively set production parameters properly, this complete database would be key. The results differ significantly, and the overall Mean Squared Error of all folds is 0.014975 (MSEs in Fold 1-5 are 0.0169, 0.0109, 0.0035, 0.0156, and 0.0280). The results of the prediction by the model are as follows:

- Oil Production Rate: 0.05196 m^3/s
- Gas Production Rate: 0.050499 m^3/s

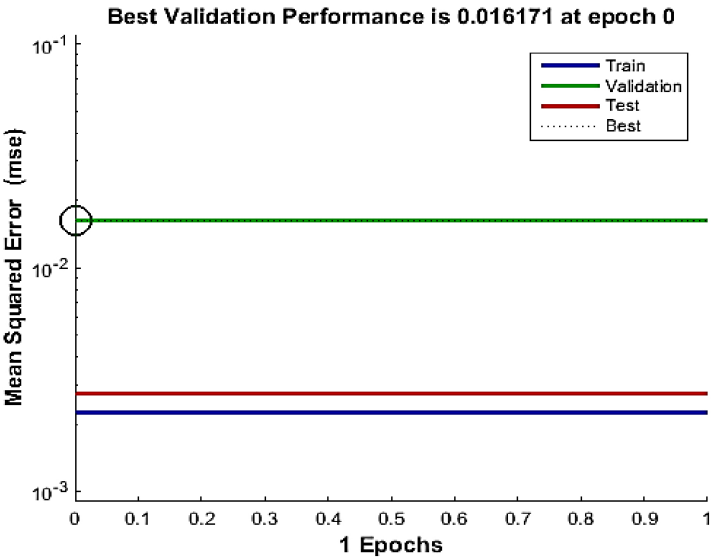

Fig. 12. Best validation permanence.

The predictive modelling process is thoroughly analysed as follows:

The validation performance of the neural network model, as depicted in Fig. 12, indicates that the model has a good initial predictive performance with a mean squared error (MSE) of 0.016171 at epoch 0. The early convergence indicates that the model quickly acquired some important patterns in the data, which reduced error at the beginning. The similarity of the values of MSE of the training and test sets also implies that the generalization of the training sets and other sets is also similar, and this is an important parameter in the assurance of reliability in application in the real world. Nevertheless, the fact that the validation error does not decrease at all after the first epoch makes one wonder whether the model itself could be further tuned (e.g., by changing learning rates or by using more complex architectures) to increase its predictive depth.

Fig. 13 offers some critical information about the practical viability of the model, which has low production rates (oil: 0.05196 m^3/s, gas: 0.050499 m^3/s) yet a high stability indicator (0.76423). This indicates that the model places more emphasis on the flow stability rather than the maximum output, which conforms to industry priorities to eliminate the hazardous conditions such as slugging or formation of hydrates. The stability versus production efficiency trade-off is also one of the main factors that should be taken into account; although the model provides a safe operation, one more way of optimization will be whether it is possible to have higher production rates without affecting the stability.

The histogram of errors in (Fig. 14) supports the trustworthiness of the model as the errors are closely concentrated around the value of zero, which means that the model is very precise in its predictions. Nonetheless, these outliers are few and point to the fact that some edge cases (e.g., extreme wellbore conditions) might need more training data or feature engineering in order to achieve better robustness.

The (Fig. 15) training dynamics indicate that the gradient descent (gradient: 0.017409 at epoch 1) remains stable without validation failure, indicating that the model was not over-fitting or diverging. The correlation plots in Fig. 16 further support this stability by showing that the R-values are close to 1, which suggests that the predicted and actual values were nearly perfectly related in a linear manner. This would show great predictive

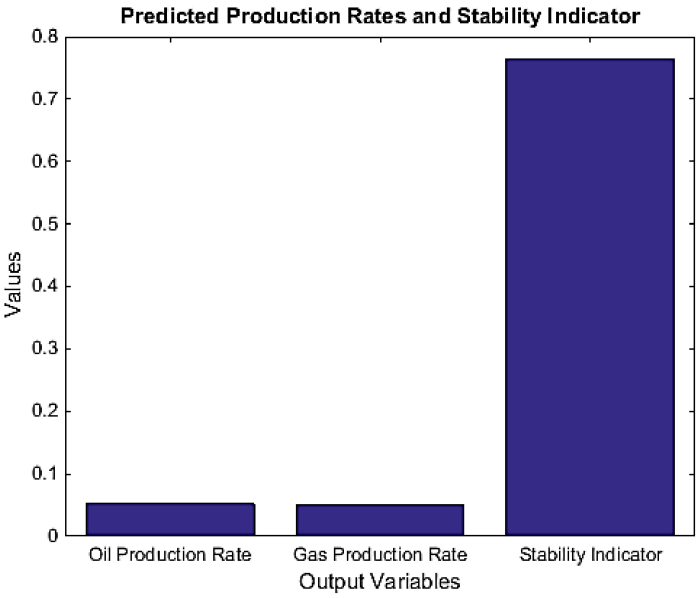

Fig. 13. Production rates and stability indicator.

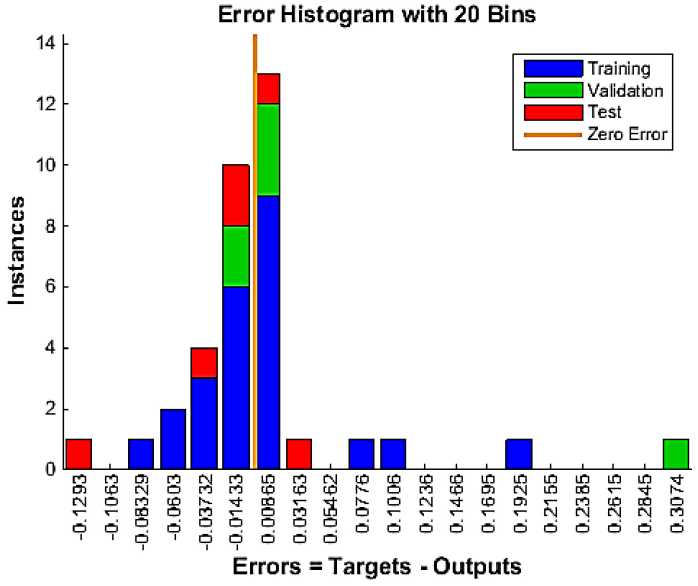

Fig. 14. Error histogram for producing parameters.

performance; the close optimality of the correlations would also be the cause of examining whether this model is over-fit to the training distribution and may fail to predict non-observable, strongly varying field data. It may be suggested that the model will be put to the test with noisy or incomplete data in the future to assess its resilience in less controlled conditions.

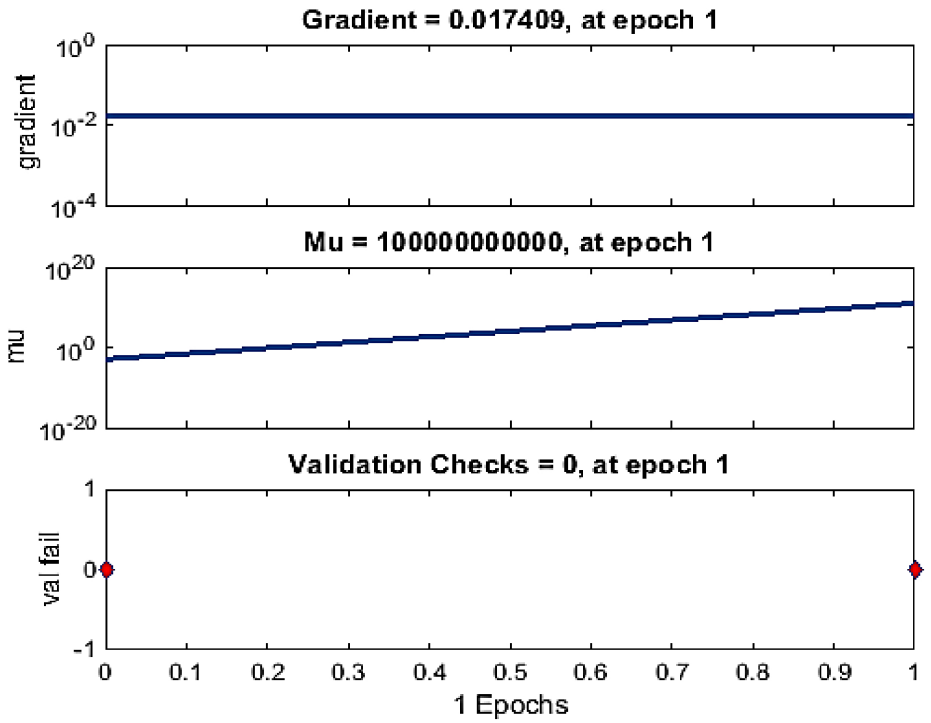

Fig. 15. Training dynamics checks of the model.

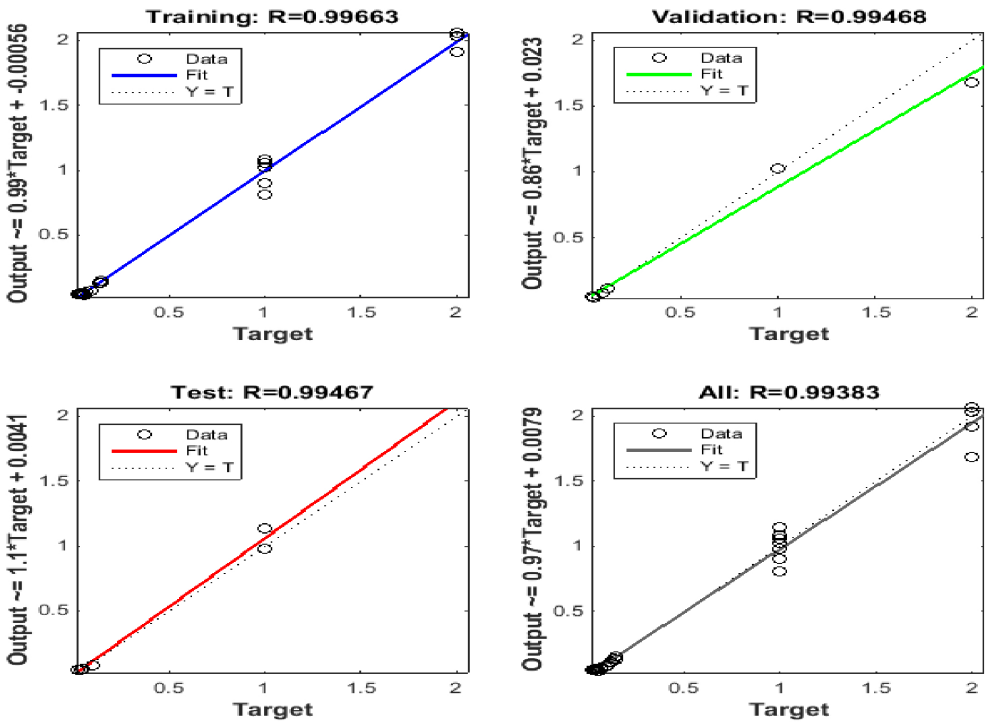

Fig. 16. Regressing test for producing parameters.

4. Conclusions

The proposed study referred to the issue of forecasting and optimization of multi-phase flow in deep wellbores, where more conventional models are not able to manage the intricate relationships between flow patterns, pressure, and temperature. The originality of the study was reflected in its combination of neural networks with mathematical modelling to categorize flow patterns (bubble, slug, annular) in addition to optimizing the production parameters, improving both efficiency and safety in hydrocarbon production. It involved a multifaceted methodology that integrated experimental simulations, data collection by sensors, and machine learning (via the Neural Network Toolbox in MATLAB) to train predictive models based on major variables that included depth, pressure, flow rates, and fluid properties. The key results of this study are explained below:

- The neural network was found to predict slug flow (100% accuracy) with added input features (e.g., viscosity, wellbore geometry), whereas only 50% accuracy was found with simple inputs (depth, pressure, flow rates).
- At the deeper levels, the preponderance was of the annular flow, and in its steep pressure gradients (which are essential to the security), whereas the bubble flow was more susceptible to changes of depth.
- The model forecasted low production (oil: 0.05196 m^3/s, gas: 0.050499 m^3/s) and high stability (indicator: 0.76423), putting the safety of operations first, and the maximum production second.
- Error histograms indicated that there were few deviations (zero errors were the most frequent), and correlation plots indicated R-values near 1, indicating the reliability of the model.
- The difference in flow patterns with ANOVA revealed significant differences between flow patterns ($p < 0.05$), which supported the strength of the classifications of the neural network.

These results showed that machine learning would be highly effective in predicting deep wellbore flow patterns and optimizing production. Since the model allows balancing the effectiveness of reservoir functions with stability, engineers are able to design wellbores in a safer and more efficient manner. Nevertheless, the model can be enhanced in subsequent research, such as reinforcement learning, to adapt the model to the alteration in the structure of the earth during mining. The results of related studies can establish a correlation of the examined theory with practical engineering and contribute to increased safety and efficiency in hydrocarbon operations.

5. Practical applications and recommendations

Consequently, the operators in the oil and gas industry will be able to optimize the process of extracting resources, operate more safely and reduce expenses with the help of the impact of data analysis on their work. Using neural networks in industrial operations allows companies to (1) reach peak productivity by adjusting the rates and settings on the spot with real-time predictions, (2) identify problems ahead of time like slugging or unstable pressure changes, avoiding equipment damage and possible dangers in deep wellbores and (3) make better choices for wellbore design, guided by the findings from the modeling. This research extends operational efficiency and forms a pathway between the theory of fluid dynamics and the field of hydrocarbon extraction. Still, to obtain the best, further studies should focus on expanding the datasets to include a wider range of

geological environments, harsh conditions and uncommon reservoirs, promoting broader usability. With reinforcement learning, control systems may be capable of self-adjusting based on wellbore changes, leading the industry on a path to autonomous operations.

Ethics declarations

Competing interests

The authors declare no competing interests.

Consent to participate

We know of no conflict of interest associated with this publication and I as corresponding author confirm that the manuscript has been read and approved for submission by all the named authors.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

Data is provided within the manuscript.

Author contributions

R.S. and F.L. wrote the main manuscript text and designed the model and the computational framework. R.S. prepared the data and wrote the MATLAB codes. F.L. suggested the methodology. M.A. and R.S. prepared figures. A.N., R.S., and M.A. discussed the results. A.N. and F.L. validated and checked the results. All authors reviewed the manuscript.

References

1. Ali AA, Abdul-Majeed GH, Al-Sarkhi A. Review of multiphase flow models in the petroleum engineering: Classifications, simulator types, and applications. Arabian Journal for Science and Engineering. 2025;50(7):4413–4456. https://doi.org/10.1007/s13369-024-09302-0.
2. Khashman MA, Shirazi H, Al-Dujaili AN. Comparative evaluation of productivity indicators in carbonate reservoir modeling by a case study for the Mishrif Formation in the Iraqi Buzurgan Oilfield. Discover Geoscience. 2025;3(1):17. https://doi.org/10.1007/s44288-025-00118-5.
3. MING F, WANG J, LIU W, *et al.* Review of multiphase flow and fluid-structure interaction of high-speed water entry. Acta Aerodynamica Sinica. 2024;42(1):68–85. https://doi.org/10.7638/kqdlxxb-2023.0197.
4. Mohammad AT, Al-Obaidi MA, Hameed EM, Basheer BN and Mujtaba IM. Modelling the chlorophenol removal from wastewater via reverse osmosis process using a multilayer artificial neural network with genetic algorithm. Journal of Water Process Engineering. 2020;33:100993. https://doi.org/10.1016/j.jwpe.2019.100993.
5. Zhao Y, Li H, Zhou H *et al.* A review of graph neural network applications in mechanics-related domains. Artif Intell Rev. 2024;57:315. https://doi.org/10.1007/s10462-024-10931-y.

6. Yadigaroglu G, Hetsroni G, Hewitt GF. Flow Regimes. In: Yadigaroglu G, Hewitt G. (eds) Introduction to Multiphase Flow. Zurich Lectures on Multiphase Flow. Springer, Cham. 2018. https://doi.org/10.1007/978-3-319-58718-9.
7. Xiao H, He S, Chen M, Liu C, Zhang Q, Zhang R. Two-Phase Production Performance of Multistage Fractured Horizontal Wells in Shale Gas Reservoir. Processes. 2025;13(2):563. https://doi.org/10.3390/pr13020563.
8. Issa MA, Issa MA, Hadi FA, Al-Zuobaidi AA, Faraj HA. Employing formation cutting removal and geomechanical models to address the pipe sticking phenomenon. Geosystem Engineering. 2025;1–17. https://doi.org/10.1080/12269328.2025.2470316.
9. Fan D, Lai S, Sun H, Yang Y, Yang C, Fan N, Wang M. Review of Machine Learning Methods for Steady State Capacity and Transient Production Forecasting in Oil and Gas Reservoir. Energies. 2025;18(4):842. https://doi.org/10.3390/en18040842.
10. Obi CE, Hasan AR, Rahman MA, Banerjee D. Multiphase flow challenges in drilling, completions, and injection: Part 1. Petroleum. 2024. https://doi.org/10.1016/j.petlm.2024.05.001.
11. Meng L, Zhang X, Jin Y, Yan K, Li Z, Fu H, Li S. Prediction model and real-time diagnostics of hydraulic fracturing pressure for highly deviated wells in deep oil and gas reservoirs. Scientific Reports. 2025;15(1):3559. https://doi.org/10.1038/s41598-025-88027-y.
12. Abdelhafiz MM, Oppelt J, Mahmoud O, *et al.* Effect of drilling and wellbore geometry parameters on wellbore temperature profile: Implications for geothermal production. Advances in Geo-Energy Research. 2023;8(3):170–180. https://doi.org/10.46690/ager.2023.06.04.
13. Skorobogatov A, Pshenin V, Tsvetkova CP, Borisov RA. Multiphase Oil-water Flow in Horizontal and Inclined Pipelines. Effect of Flow Velocity on Flow Patterns. International Journal of Engineering. 2025;38:1820–1830. 10.5829/ije.2025.38.08b.08.
14. Raissi M, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics. 2020;378:686–707. https://doi.org/10.1016/j.jcp.2018.10.045.
15. Yang H, Li J, Zhang G, Zhang H, Guo B, Chen W. Wellbore multiphase flow behaviors during gas invasion in deepwater downhole dual-gradient drilling based on oil-based drilling fluid. Energy Reports. 2022;8:2843–2858. https://doi.org/10.1016/j.egyr.2022.01.244.
16. Nakayama H, Yamasaki Y, Ohashi K, Nakaya S. Simple method for estimating well yield potential through geostatistical approach in fractured crystalline rock formations. Journal of Hydrology. 2021;597:125719. ISSN 0022-1694. https://doi.org/10.1016/j.jhydrol.2020.125719.
17. Feng Q, Li R, Jia Y, Liang Z. Big data and artificial intelligence-based optimization of petroleum exploration and reservoir modeling: Intelligent pathways for enhancing efficiency and accuracy. Advances in Resources Research. 2025;5(2):477–495. https://doi.org/10.50908/arr.5.2_477.
18. Carrillo F, Soulaine C, Bourg I. Multiphase Flow Modeling in Multiscale Porous Media: An Open-Source Micro-Continuum Approach. Journal of Computational Physics X. 2020;8. 10.1016/j.jcpx.2020.100073.
19. Mejail M, Nestares BK. Mathematical Modeling of Multi-Phase Flow in Porous Media for Environmental and Petroleum Engineering. Journal of Applied Mathematical Models in Engineering. 2025;1(3):1–7. https://theeducationjournals.com/index.php/JAMME/article/view/202.
20. Jiang J. Simulating multiphase flow in fractured media with graph neural networks. Physics of Fluids. 2024;36(2). https://doi.org/10.1063/5.0189174.
21. Zhu LT, Chen XZ, Ouyang B, Yan WC, Lei H, Chen Z, Luo ZH. Review of machine learning for hydrodynamics, transport, and reactions in multiphase flows and reactors. Industrial & Engineering Chemistry Research. 2022;61(28):9901–9949. https://doi.org/10.1021/acs.iecr.2c01036.
22. Al-Naser M, Elshafei M, Al-sarkhi A. Artificial neural network application for multiphase flow patterns detection: A new approach. Journal of Petroleum Science and Engineering. 2016;145:548–564. DOI:10.1016/J.PETROL.2016.06.029.
23. Mount N, Maier H, Toth E, Elshorbagy A, Solomatine D, Chang F-J, Abrahart RJ. Data-driven modelling approaches for socio-hydrology: Opportunities and challenges within the Panta Rhei Science Plan. Hydrological Sciences Journal. 2016;61. DOI:10.1080/02626667.2016.1159683.
24. Ekong AP, James GG, Ohaeri I. Oil and gas pipeline leakage detection using iot and deep learning algorithm. Journal of Information Systems and Informatics. 2024;6(1):421–434. https://doi.org/10.51519/journalisi.v6i1.652.
25. Mercante R, Netto T. Virtual Meter with Flow Pattern Recognition Using Deep Learning Neural Networks: Experiments and Analyses. SPE Journal. 2024;29:2181–2196. 10.2118/219465-PA.
26. Hamed U-A, Sadeq DJ, Mahmud HB. Multiphase Flow Behavior Prediction and Optimal Correlation Selection for Vertical Lift Performance in Faihaa Oil Field, Iraq. Iraqi Journal of Chemical and Petroleum Engineering. 2023;24(4):127–140. https://doi.org/10.31699/IJCPE.2023.4.13.

27. Yuan J, Zhang X. Multiphase Flow Simulation Using Machine Learning: A Review. Journal of Computational Physics. 2022;454:110920. https://doi.org/10.1016/j.jcp.2022.110920.
28. Vardy AE, Tijsseling AS. On Boussinesq and Coriolis coefficients and their derivatives in transient pipe flows. Applied Mathematical Modelling. 2022;111:454–470, ISSN 0307-904X. https://doi.org/10.1016/j.apm.2022.06.029.
29. Omari A, Wang C, Li Y, Xu X. The progress of enhanced gas recovery (EGR) in shale gas reservoirs: A review of theory, experiments, and simulations. Journal of Petroleum Science and Engineering. 2022;213:110461, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2022.110461.
30. Shahade AK, Walse KH, Thakare VM. Deep learning approach-based hybrid fine-tuned Smith algorithm with Adam optimiser for multilingual opinion mining. International Journal of Computer Applications in Technology. 2023;73(1):50–65. https://doi.org/10.1504/IJCAT.2023.134080.
31. Steckler S, Orescanin M, Ortiz P, Petkovic V. Beyond Mean Squared Error: Alternative Loss Functions for Geoscience Image Regression with Neural Networks. In IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. 2024 July;7359–7362. IEEE. https://doi.org/10.1109/IGARSS53475.2024.10641070.