ORIGINAL ARTICLE

Two-phase oil and water flow pattern identification in vertical pipes applying long short-term memory networks

Carlos Mauricio Ruiz-Díaz[1] · Brayan Quispe-Suarez[2] · Octavio Andrés González-Estrada[2]

Abstract

Accurate biphasic flow pattern recognition is essential in the design of coatings for the oil and gas sector because it enables engineers to create materials that are tailored to specific flow conditions. This results in enhanced corrosion protection, erosion resistance, flow efficiency, and overall performance of equipment and infrastructure in the challenging environments of the oil and gas industry. The development of flow maps has been based on empirical correlations that incorporate characteristics such as superficial velocities, volume fractions, and physical properties such as the density and viscosity of the analyzed substances. In addition, geometric parameters such as the inclination and the internal diameter of the pipes are considered. However, due to the difficult working conditions on offshore platforms and the limitations in monitoring internal flow patterns, technological advances have been implemented to improve this process using artificial intelligence techniques. In this context, this study proposes using a long-term memory (LSTM) recurrent neural network to predict the flow patterns generated in vertical pipes. This LSTM network was trained and validated using data obtained from a literature database. The results obtained showed that the model has a prediction error of less than 1%. These technological advances represent an important step towards optimizing the flow pattern identification process in the hydrocarbons industry. By leveraging the capabilities of artificial intelligence, more accurate and reliable forecasts can be obtained, enabling informed decisions and improving the efficiency and safety of operations.

Keywords LSTM · Recurrent neural networks · Liquid–liquid two-phase flow · Flow pattern identification

1 Introduction

Multiphase flow transport must be considered in designing and optimizing many engineering systems, such as food processing equipment, fuel cells, and the petrochemical industry [1]. The latter has focused its current research on crude oil transport [2]. The optimization of production in oil fields occurs with the continuous monitoring of flow rates, temperature, and pressure of the fluid extracted from wells. Performance can be measured against extraction costs, which is possible with reliable data that provide a clear perspective on fluid transport. Based on the abovementioned concepts, studies have been developed to model multiphase flows in transport lines worldwide [3]–[4].

Predicting flow patterns in the oil and gas sector is essential to optimize operational efficiency, maximize production, ensure safety, comply with regulations, prevent equipment wear, and minimize environmental impact. It also facilitates the implementation of predictive maintenance strategies and contributes to continuous and reliable operation while reducing costs and improving risk management in a critical industry [5]. Coatings and thin films are often applied to surfaces in the oil and gas industry to protect equipment and infrastructure from corrosion caused by contact with corrosive fluids [6]. Understanding the biphasic flow patterns of these fluids can help design coatings that are optimized for specific flow conditions, ensuring effective corrosion protection. Moreover, solid particles suspended in the fluid can cause erosion of surfaces over time. Recognizing flow patterns helps design coatings that are resistant to erosive wear, extending the lifespan of equipment and reducing the

✉ Octavio Andrés González-Estrada
agonzale@uis.edu.co

1 Industrial Multiphase Flow Laboratory (LEMI), São Carlos School of Engineering, University of São Paulo, São Carlos 13566-590, Brazil

2 Research Group in Energy and Environment (GIEMA), Universidad Industrial de Santander, Cr 27 C 9, Bucaramanga 680002, Colombia

جامعة قطر
QATAR UNIVERSITY

🅐 Springer

need for frequent maintenance [7]. Certain flow patterns can lead to the accumulation of hydrates and waxes, which can obstruct pipelines and impede fluid flow. Coatings and thin films can be engineered to inhibit the adherence of these substances to surfaces, maintaining efficient flow even under challenging conditions.

The studies developed by [8] demonstrate the predictive ability of artificial neural networks fed by a class of data-driven virtual flow meters (VFM). This work showed that neural networks could be a promising tool for predicting multiphase flow rates using pressure and temperature data. Conversely, as indicated by the research presented in reference [9], the update frequency of a steady-state VFM model is critical for sustaining optimal performance over time, particularly in dynamic, non-stationary conditions.

The study of multiphase flows integrates concepts directly related to transport phenomena in modeling and flow characterization, using the principles of conservation of mass and conservation of momentum. These principles allow for a mathematical approximation of flow characteristics inside pipelines [10]. The description of multiphase flow behavior requires the use of experimental data or correlations due to the complexity involved in their study. Although advanced computational methods are being developed, data and correlations are generally still needed to implement multiphase flow modeling effectively [11].

Multiphase flows can be found in multiple engineering practices, especially in the oil industry, where two-phase and tri-phase flows can be found [12]. Two-phase flow refers to the interactive flow of two distinct phases (each phase representing a mass or volume of matter) with common interfaces in a channel. Possible combinations are solid–liquid, where solid particles are dispersed primarily in liquid; solid–gas, where a gas stream transports solid particles; and liquid–vapor (gas), where the volume fraction of one phase relative to the other results in different flow regimes and combination of the above [13].

At present, there is a large number of studies on the behavior of multiphase flows. One approach being explored is data-driven modeling, a technique that analyzes system data and looks for relationships between input and output variables without relying on knowledge of the underlying physical equations [14, 15]. This promising approach makes it possible to take advantage of the information contained in the data and discover patterns and correlations that can be used to understand and predict the behavior of multiphase flows efficiently and accurately. This poses challenges when supplying input data; thus, it is essential to give adequate statistical treatment and avoid a poor fit in model parameters [16]. This process is achieved by implementing artificial neural networks, which start by calculating a function with the inputs given, propagating the calculated values from the input neurons to the output neurons and using the weights as intermediate parameters. Learning occurs by changing the weights that connect the neurons [17] to have predictions that fit reality.

The study in [18] used daily data on the production of oil, water, and gas generated from an oil field as the primary input variable to model a neural network, resulting in predictions of multiphase production in the short term. The long short-term memory (LSTM) algorithm was implemented to generate the model. This approach showed that, compared to the model based on numerical simulation physics, this neural network approach was less dependent on engineering calculations and offered faster and more robust solutions. In [19], an investigation was conducted using artificial intelligence systems with actual data from an oil platform in the South China Sea to develop a virtual flow meter model. This work demonstrated how LSTM considers the influence of previous time steps on the evaluated flow rate, capturing the relationship between parameters such as temperature and pressure on the predicted flow rate. LSTM is also highlighted in [20], where it was shown that precision is improved if the input and output sequence lengths are equal and how to remove this limitation without losing precision in the neural network.

This work presents the development of a recurrent neural network model implementing the LSTM algorithm to predict flow patterns generated within vertical two-phase liquid–liquid flow pipes of lubricating oil and water. First, two-phase flow data is collected and processed. The following section discusses structuring the LSTM neural network, followed by the evaluation and optimization approaches. Finally, results are presented for different network configurations, followed by concluding remarks.

2 Methodology

2.1 Data processing

The data used in this study was collected by [21], who compiled relevant information on the two-phase flow of oil and water in a vertical pipe. For this, the works of 18 authors previously published in the literature were analyzed. These collected data provide a solid and representative basis for the development of this work, allowing an exhaustive and well-founded analysis of the behavior of the biphasic flow in this specific configuration. Table 1 presents the information in the database, summarizing the quantitative characteristics of each analyzed work.

Different types of two-phase flows were considered in the study, including dispersed oil-in-water (D O/W), dispersed water-in-oil (D W/O), churn oil-in-water (churn O/W), churn water-in-oil (churn W/O), core flow, stratified oil-in-water (S O/W), stratified water-in-oil (S W/O), flow in transition

Table 1 Summary of the collected database from reference authors

Author	D[m]	μ_o[Pa.s]	ρ_o[kg/m^3]	N°. flow patterns	N°. data
[3]	0.0284	0.488	925	3	392
[22]	0.1064	0.00188	800	6	49
[23]	0.02	0.011	856	3	101
[24]	0.05	0.02	850	5	126
[25]	0.04	0.035	860	4	92
[26]	0.0263	0.0201	851	2	15
[27]	0.02	0.125	801	4	175
[28]	0.127	0.001	801	2	9
[8]	0.027	0.02	801	4	109
[29]	0.0254	0.001	792	4	55
[30]	0.0254	0.001	792	5	32
[31]	0.01	0.287	1823	3	1421
[32]	0.026	0.0011	793	5	90
[33]	0.03	0.03	856	4	252
[34]	0.0284	0.5	930	1	92
[35]	0.03	0.044	860	4	98
[36]	0.02	0.09732	857	7	1699
[37]	0.04	0.0041	824	1	57

(TF), very fine dispersed oil-in-water (VFD O/W), and very fine dispersed water-in-oil (VFD W/O); see Fig. 1. These different types of two-phase flows represent various configurations and flow patterns that can occur in oil and water systems and were considered in order to better understand the characteristics and behavior of multiphase flows.

The variables used to build the database are the superficial velocities of the water phase (j_w) and oil phase (j_o). These velocities are function of injection rates and the cross-sectional area of the pipe, as well as the mixture velocity (J) and the volumetric fractions of oil–water fluids (ε_o and ε_w) that characterize the flow. Other important parameters included in the data collection are the oil viscosity (μ_o), oil density (ρ_o), and the internal diameter of the pipe (D). Specific geometrical configurations are generated for each combination of parameters, grouped for this work into the ten flow patterns, which are the output data of the intelligent model to be developed. Table 2 presents the specific amount of data available for each flow pattern.

2.1.1 Normalization

To train the neural network, it is crucial to normalize the information that will be used as inputs to the intelligent model, providing precise data that allows the network to relate the supplied variables with the predicted pattern. Normalization is defined as a scale change of the original data such that it falls within a specific range [38]. The input vector of the network is normalized, which contains $x = [j_o, j_w, j, \varepsilon_o, \varepsilon_w, D, \mu_o]$. Equation (1) presents the mathematical definition used to normalize the numerical information:

$$X_{\text{normalized}} = \frac{x - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

The variable $X_{\max}$ represents the maximum value of the series under study, while $X_{\min}$ represents the minimum value of the series, and x corresponds to the point value to be normalized.

Fig. 1 Two-phase flow patterns in vertical pipes
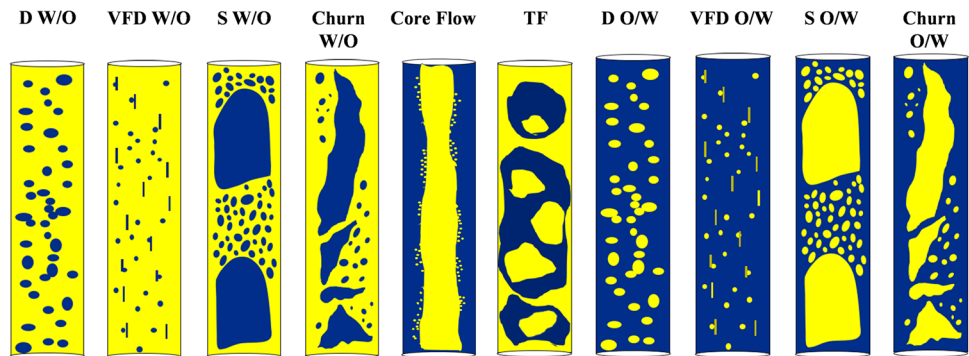
Table 2 Quantification of data for each flow pattern

Flow pattern type	D O/W	D W/O	Churn O/W	Churn W/O	Core flow	S O/W	S W/O	TF	VFD O/W	VFD W/O
Amount of data	947	1292	21	127	481	460	657	315	366	205
% of total	19%	27%	0.40%	3%	10%	9%	13%	6%	8%	4%

2.1.2 Codification

Given that the flow pattern is defined in the database as a categorical variable, it is necessary to encode the label that is assigned to the ten different configurations. To carry out this encoding, a number from 0 to 9 is assigned to each of the defined flow patterns. Since there are different data labels, simple encoding is insufficient. In this case, encoding is achieved using the OneHotEncoder function, a well-established technique commonly employed in the fields of data science and machine learning. The fundamental concept behind one-hot encoding is to represent categorical variables as binary vectors. Each category is transformed into a separate column with binary values (1 or 0) to indicate the presence or absence of that category in a data row. This process eliminates the encoded variable as a whole and introduces a binary variable for each unique integer value. Table 3 shows the result of encoding the ten flow patterns with their respective binary representation.

2.2 LSTM model structuring

The identification of two-phase flow patterns in vertical oil pipelines is an inherently sequential and time-dependent problem. Flow patterns change in response to previous events and evolve continuously. In this context, the use of long short-term memory networks (LSTM) is justified due to its ability to capture sequential dependencies and handle time series data. While support vector machines (SVM), artificial neural networks (ANN), K-means, K-nearest neighbors (KNN), and similar techniques are highly effective in classification and clustering problems, they do not possess the inherent capacity of LSTM to effectively model and predict data sequences. With all the processed information from the reference database, we proceed with the definition of the internal structure of the LSTM. A general scheme of the structure of this type of neural network is presented in Fig. 2, as well as the vector with the input parameters that will be used to generate the predictions. Figure 2 shows how the hidden layers are connected sequentially to process the

Table 3 OneHotEncoder codification for the flow patterns

ID	Flow pattern	OneHotEncoder application									
0	D O/W	1	0	0	0	0	0	0	0	0	0
1	D W/O	0	1	0	0	0	0	0	0	0	0
2	Churn O/W	0	0	1	0	0	0	0	0	0	0
3	Churn W/O	0	0	0	1	0	0	0	0	0	0
4	CoreFlow	0	0	0	0	1	0	0	0	0	0
5	S O/W	0	0	0	0	0	1	0	0	0	0
6	S W/O	0	0	0	0	0	0	1	0	0	0
7	TF	0	0	0	0	0	0	0	1	0	0
8	VFD O/W	0	0	0	0	0	0	0	0	1	0
9	VFD W/O	0	0	0	0	0	0	0	0	0	1

Fig. 2 General structure of an LSTM with input, hidden, and output layers
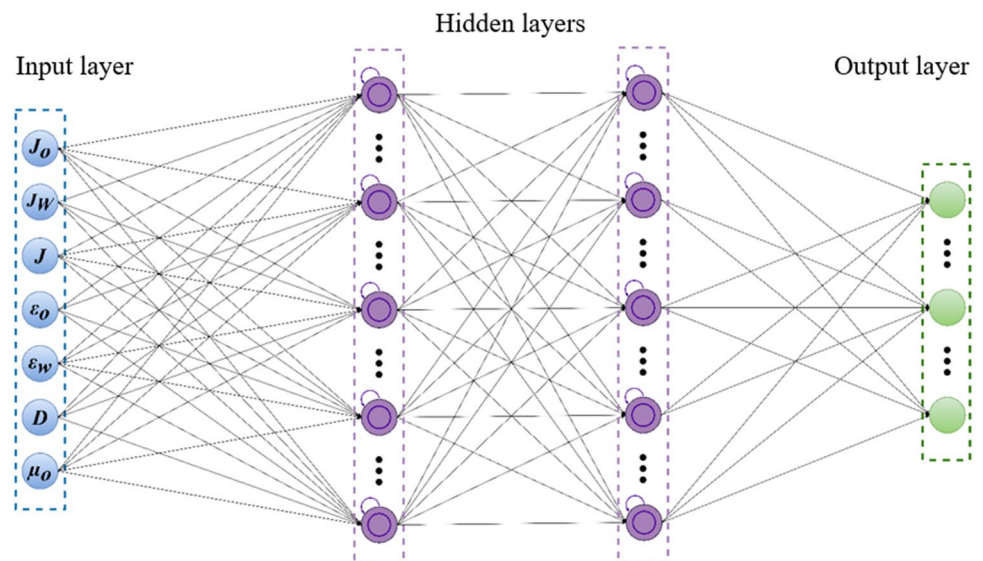
data from inputs layers. In LSTM, the combination of the current input and the previous hidden state, along with the LSTM's various gates (including the forget gate, input gate, and output gate), determines how the cell state is updated and what information is retained or forgotten in the memory. The input vector could be interpreted as a three-dimensional array, composed of samples, time series, and features. Once the input vector with the parameters is defined, the categorical representation is generated, which will be the output of the predictive model. In this case, the output has been assigned the flow pattern developed inside the vertical pipes.

The operation of an LSTM is based on the internal process that occurs in its state cells. These cells integrate activation functions to process both new information and information from previous state cells, thus allowing sequential processing and retention of relevant information over time. This mechanism of integration and processing of information in the state cells is essential for the effective operation of the LSTM. Figure 3 presents a diagram of the logical sequence that a state cell follows internally. Several steps occur before obtaining a hidden state or an output from the cell, which involve multiple transformations of the data.

The logical sequence developed within the LSTM cells allows for a detailed explanation of the order in which information is processed within them. Thus, Fig. 3a presents a filter for which information is retained and which is discarded in the cell state. This operation is carried out thanks to the sigmoid layer (σ) known as the *forget gate*:

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{2}$$

This gate analyzes h_{t-1} and X_t, which correspond to the current input and the output of the previous state, and generates a number between 0 and 1 for each number in the cell state C_{t-1}. W is the weighting function adjusted inside the algorithm during the training and b is the bias. Thus, when the function generates a 1, it saves this information, while when it generates a value of 0, it skips the processed information.

Then, the neural network selects the information that will be stored in the cell state (Fig. 3b). This processing is done in two parts: in the first part, a sigmoid layer evaluates the values to be updated in the input gate i_t, and in the second part, a layer with the hyperbolic tangent function (*tanh*) creates a new vector with candidate values C_t, some of which
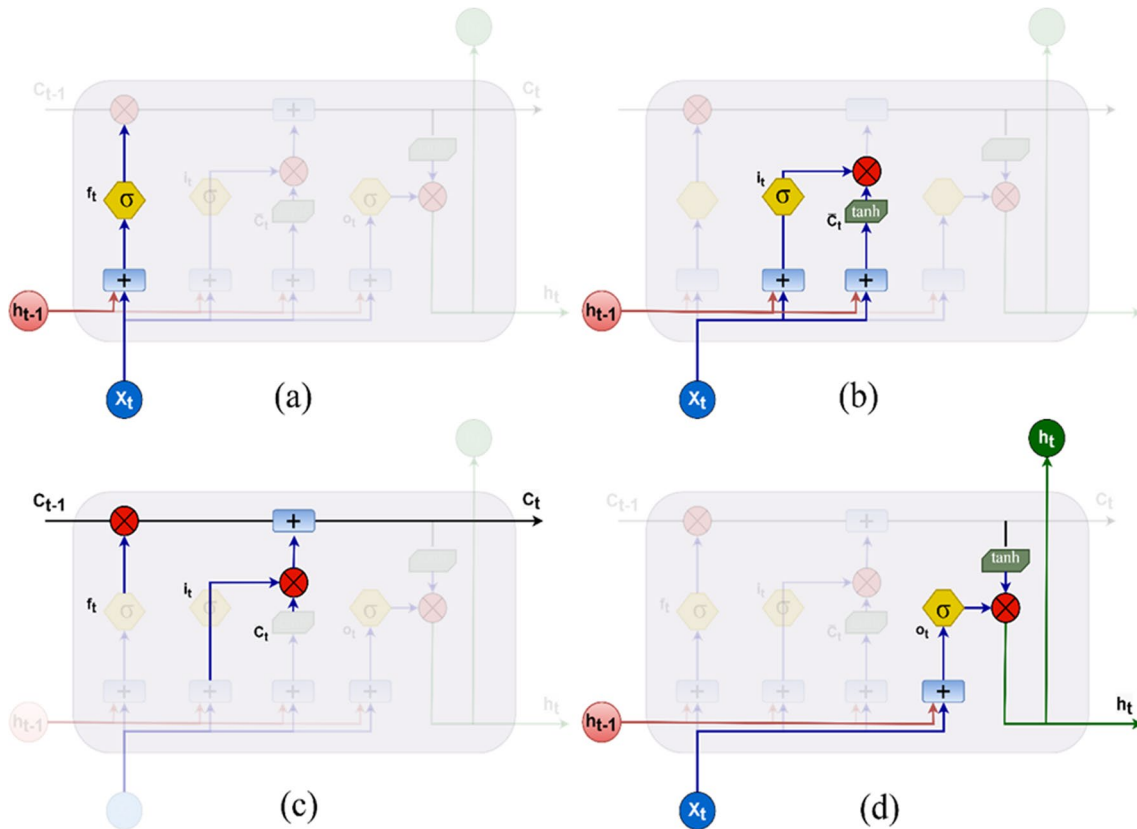

Fig. 3 Logic sequence developed inside the LSTM state cell: **a** first logic order, **b** second logic order, **c** third logic order, **d** completion of the sequence. Sigmoid layer σ, hyperbolic tangent function tanh, input h_{t-1}, and output X_t of the previous state. The hidden layers produce the current hidden state h_t as the output

will be added to the cell state. This sequence is synthesized in Eqs. (3) and (4):

$$i_t = \sigma\left(W_i[h_{t-1}, x_t] + b_i\right) \tag{3}$$

$$\widetilde{C}_t = tanh\left(W_c[h_{t-1}, x_t] + b_c\right) \tag{4}$$

After selecting the information to be included in the cell state, the cell state is updated (Fig. 3c). The previous state, C_{t-1}, is converted to $\widetilde{C}_t$ by multiplying it with f_t, and then adding $i_t C_t$. This results in Eq. (5):

$$C_t = f_t C_{t-1} + i_t C_t \tag{5}$$

Finally, in Fig. 3d, the value that will continue the process to the upper cells is selected. The output gate o_t is based on the state of the cell, and a sigmoid layer is executed to decide which parts of the cell state are generated. Then, the cell state is processed with a hyperbolic tangent function that converts the values to a range between -1 and 1, which is multiplied by the output of the sigmoid gate such that only the selected parts are endorsed, generating Eqs. (6) and (7):

$$o_t = \sigma\left(W_o \bullet [h_{t-1}, x_t] + b_o\right) \tag{6}$$

$$h_t = o_t tanh(C_t) \tag{7}$$

In problem-solving using neural networks, objective functions tend to be nonlinear. The nature of artificial neurons is logistic regression, which results in a linear function:

$$P = \frac{1}{1 + e^{-y}} \tag{8}$$

$$y = a_1 x_1 + \cdots + a_n x_n \tag{9}$$

Equation (8) shows the logistic regression function, which is a discrete binary function, while Eq. (9) represents a linear function. When multiple neurons are grouped into multiple layers, the network becomes a super neuron, whose goal is a linear function. In this way, activation functions are implemented for neural networks, which generate non-linearity in the objective function. Neural networks use activation functions that are nonlinear to develop the learning process [39].

2.3 *Model evaluation and optimization*

The loss function is essential to evaluate the performance of an intelligent model. In this context, key statistical parameters are used to determine the performance and accuracy of the proposed models. The mean square error (MSE) and the cross-entropy are two of the statistical parameters used for this purpose. These parameters allow us to quantify and measure the discrepancy between the model predictions and

the actual values, providing a quantitative assessment of the accuracy and quality of the model predictions. The mean square error reads

$$MSE = \frac{\sum_{i=1}^{n}\left(y_i - \lambda(x_i)\right)^2}{n} \tag{10}$$

where y_i is the target value for the test instance x_i, $\lambda(x_i)$ is the expected target value for the test instance x_i, and n is the number of test instances.

For cross-entropy, H is the cross-entropy function, P is the target distribution, and Q is the approximation of the target distribution. Cross-entropy is calculated using probabilities of events from P and Q as

$$H(P, Q) = -\sum_{x \in X} P(x) \times log(Q(x)) \tag{11}$$

In order to optimize the final model, taking into account its evaluation, different configurations are defined by varying parameters of the neural network structure, such as the number of cells per layer, the number of stacked hidden layers, the activation function of the model, and the loss function. All these adjustments are made using open-source code in Python v3.9. The parameters for each type of LSTM configuration are presented in Table 4.

During the structuring and development phases of the predictive model, it is crucial to pay attention to the compilation stage, where all the parameters selected to train the network are carefully considered. In the particular case of the LSTM 3 network, the model is compiled specifying the Adam optimizer and the categorical cross-entropy loss function. This configuration is implemented with the aim of achieving better performance in the classification of the series, taking into account the type of information that is being used.

The model is adjusted by updating the weights for each parameter included in the training set, considering that previous cell states are integrated into each of them. All LSTM structures were trained with 80% of the collected data. In the evaluation, the remaining 20% of the data is used to verify the results in each of the metrics specified at the time of compiling the model. A graphical analysis is carried out in Fig. 4 with the aim of selecting the model with the best performance. The values of the loss function for each model are examined both in the training phase and in the validation phase, considering a total of 300 epochs. Through this analysis, we seek to identify the model that presents a lower loss and, therefore, a better fit to the data and a greater predictive capacity.

After evaluating the performance of the different configurations, notice that during training, LSTM 1 tends to decrease the error faster than the other models, making it the optimal configuration for predicting flow patterns. Figure 4a

Table 4 Parameter definition for the structured LSTM configurations

Algorithms	Hyperparameter	Selection	Other parameters	Selection
LSTM 1	Learning rate	0.001	Activation function	Sigmoid
	Number of hidden layers	2	Loss function	Categorical cross-entropy
	Number of cells per layer	80	Optimizer	Adam
	Batch size	128	Epochs	300
LSTM 2	Learning rate	0.001	Activation function	Sigmoid
	Number of hidden layers	2	Loss function	Categorical cross-entropy
	Number of cells per layer	30	Optimizer	Adam
	Batch size	32	Epochs	300
LSTM 3	Learning rate	0.001	Activation function	Sigmoid
	Number of hidden layers	3	Loss function	Categorical cross-entropy
	Number of cells per layer	30	Optimizer	Adam
	Batch size	128	Epochs	300
LSTM 4	Learning rate	0.001	Activation function	Relu
	Number of hidden layers	2	Loss function	MSE
	Number of cells per layer	80	Optimizer	Adam
	Batch size	1	Epochs	300

Fig. 4 Evaluation of models with respect to the loss function: **a** training phase and **b** validation phase
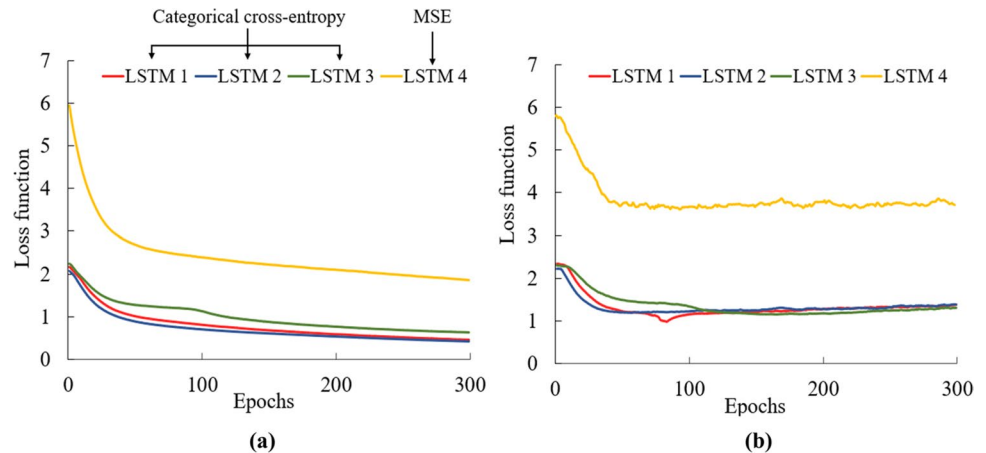
shows the performance of each of the four models on the training data, with LSTM 1 being selected over the others due to its performance in the validation phase, having the lowest error of all models. On the other hand, LSTM 4 implements a different loss equation from the others, starting its training with a considerable error and taking longer to decrease it. This is due to the characteristics of the loss function, which does not binary classify the expected flow patterns, but treats them as an integer encoding. When compared to the expected output, it creates a discrepancy in the output of the network. Figure 4b shows the performance of the networks on the validation data, detailing the behavior of the networks when faced with unknown data. The error fluctuates compared to that obtained in the training data. Values tend to converge if the loss function is the same. However, when testing with the MSE loss function, it presents an error greater than 2%, demonstrating that it is the model with the lowest performance.

Figure 5a presents the accuracy values obtained for each model during the training. LSTM models 1, 2, and 3 tend to increase their accuracy. Only LSTM model 4, which has a different loss function, presents a linear behavior stagnating at an accuracy of less than 0.1. This behavior is due to the complexity of the prediction, and this complexity is given by the need to predict one flow pattern within ten different types. With OneHotEncoder encoding, better performance is obtained than with normal encoding, due to OneHotEncoder encoding the data in a binary format. More data is needed to increase the accuracy of the network. On the other hand, Table 2 shows how a certain flow pattern represents less than 1% of the total data. This may incur in overfitting of the network, causing the model to remember the supplied data, but be inaccurate with model evaluation data.

The adjustment of the number of epochs tends to be irrelevant if it is prolonged over time, due to its convergent trend. The LSTM 1 model was selected to test the generation

Springer

Fig. 5 Evaluation of model accuracy: **a** training phase and **b** validation phase

(a)

(b)
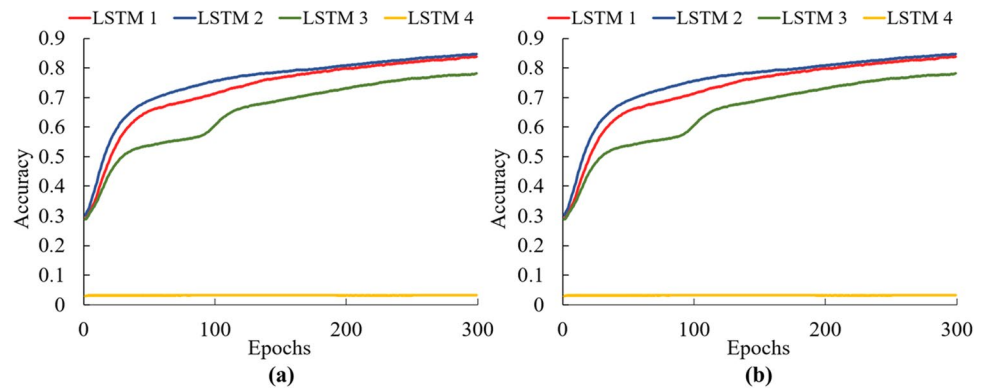
of flow maps, because the error obtained in this model is less than 1%, and its accuracy grows exponentially with the epochs, reaching a value close to 0.9 in the training. In this model, the supplied data was parameterized to observe the behavior of the model.

3 Results

In this section, the performance of the LSTM 1 model in predicting flow patterns is evaluated, and the error seen in the previous section is verified. Different flow maps generated by the LSTM 1 network using reference data are presented. After selecting the model with the best performance, in this case, LSTM 1, a flow map generation test was carried out. For this, 300 data sets were provided as input to the network. Figure 6 shows the number of times the model correctly predicted the flow pattern. Yellow marks represent the flow pattern of the provided data, while blue marks indicate the predictions generated by the network. An error of only 0.3% is observed, which agrees with the results presented in Fig. 4a. The loss function used to evaluate the model demonstrates

excellent classification performance, indicating that the model accurately fits the actual data. In this case, there was a mismatch in a single prediction, and as the amount of data increases, the error is expected to increase to 0.8% when all data is considered. Although the data provided to generate this graph only includes 8 of the 10 possible flow patterns, it is still a significant sample due to the amount of data used. The model, however, incorrectly predicted a pattern for that specific data.

The model did not generate additional patterns beyond the eight present in the range. For the specific data points in the input vector $x = [j_o, j_w, j, \varepsilon_o, \varepsilon_w, D, \mu_o]$, the model predicted a pattern of D W/O, while the expected pattern was VFD W/O. However, this is within the expected error margins.

The flow map generated with the complete dataset from reference data is presented in Fig. 7, which shows the flow pattern depending on J_w and J_o. in logarithmic scale. The flow map indicates that the decrease in the magnitudes of water and oil velocities induces the formation of the flow pattern of S W/O and D W/O. Figure 7 also shows that the increase in the magnitude of J_w and J_o generates the formation of the other types of flow patterns. Notice that the S

Fig. 6 Evaluation of the algorithm LSTM 1 for the different flow patterns
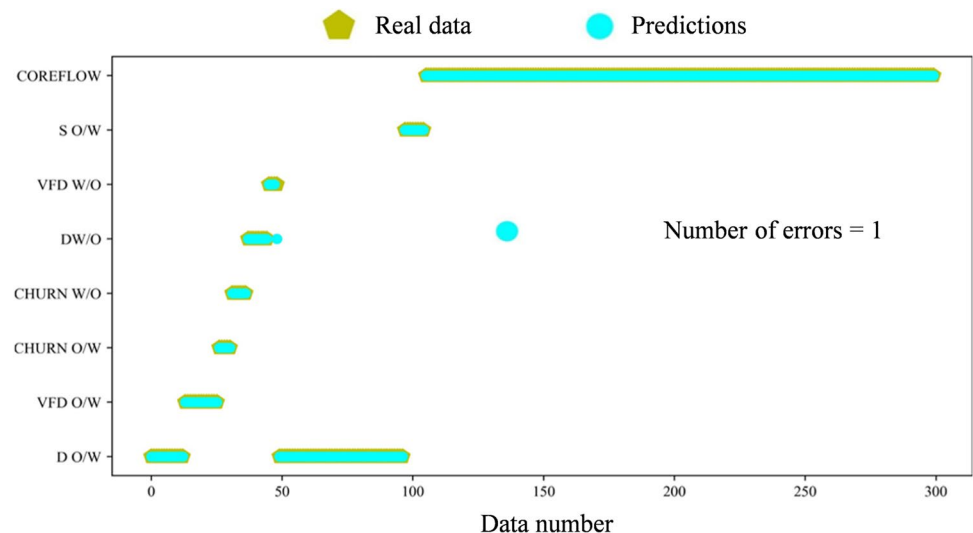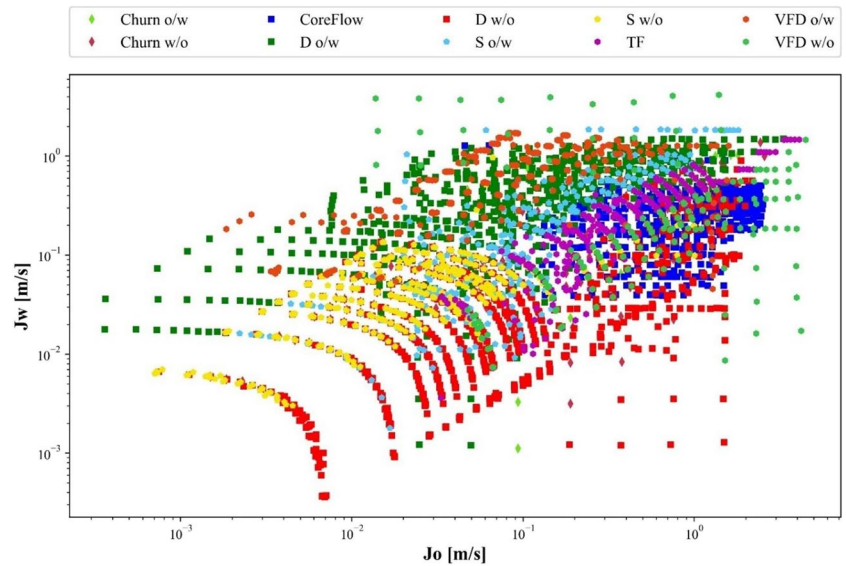
Fig. 7 Flow map with data from reference values in Table 1

O/W flow pattern presents overlapping points, which represent the influence of other variables in the formation of a flow pattern, such as internal diameter and oil viscosity.

Figure 8 shows the flow patterns obtained by implementing LSTM 1 network. The differences between the maps presented in Fig. 7 and Fig. 8 can be seen in the region of the flow map where the transition between flow patterns occurs, generated when the oil velocity is high and the water velocity is low. The magnitude ranges of the superficial velocities of the substances are 0.01 to 1 m/s for J_w and 0.1 to 1 m/s for J_o.

In this study, two flow maps were developed considering the information provided by each corresponding author. Figure 9 presents the flow map structured with the information presented by [24], in which five flow patterns (CoreFlow, D

O/W, D W/O, TF, VFD O/W) and a total of 125 data points are identified. Additionally, the separation zone between the D W/O and D O/W flow patterns is presented. Considering the magnitudes of the velocities developed by the substances, 60% of the data points are in the D O/W flow pattern. On the other hand, the CoreFlow pattern is presented in the zone of high superficial velocities for water and low velocities for oil.

The implementation process of the LSTM 1 network resulted in the flow map presented in Fig. 10. The performance of the network is low in the fluid transition zone, showing multiple prediction errors and generating patterns that are not defined in that zone. In this case, there are nine correct predictions out of 18, which is due to the behavior of the fluid in the transition area. The transition involves the

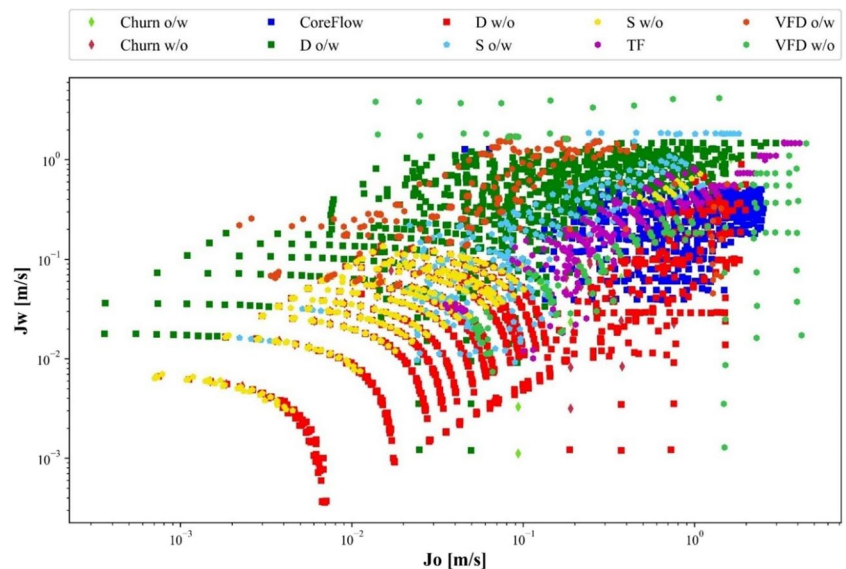**Fig. 8** Flow map generated from LSTM 1 algorithm

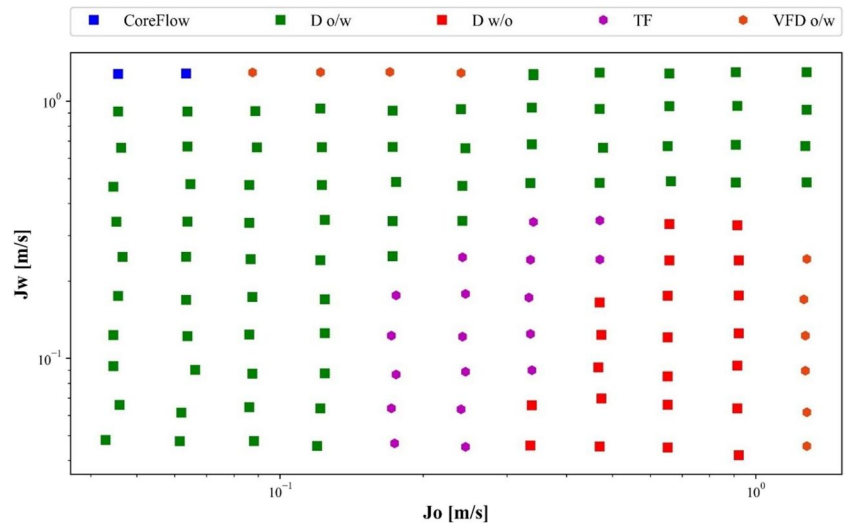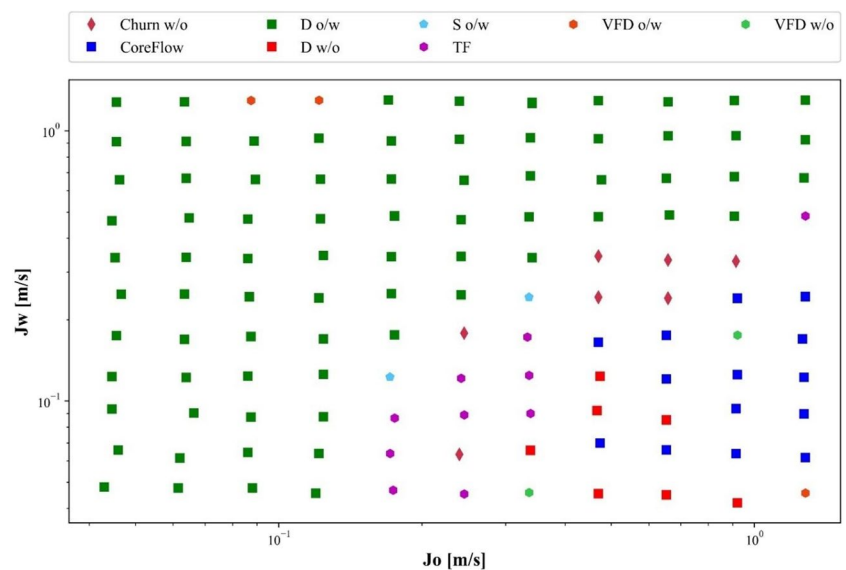**Fig. 9** Flow map with data from Flores et al. [24]

Fig. 10 Flow map generated from Flores et al. [24], applying the LSTM 1 model

change of the dominant fluid, and different authors have limited information about this zone due to its chaotic behavior, which cannot be fully controlled. However, it is worth noting that the LSTM 1 network accurately predicts the flow patterns developed in the other zones of the flow map.

Additionally, Fig. 11 presents the flow map developed using the information from [33], which shows a greater number of points in the transition zone, and also presents fewer flow patterns (D O/W, D W/O, S O/W, TF). Figure 12 shows the results obtained by processing the data with the LSTM 1 network, in which 31 correct flow patterns out of the 38 expected are identified, showing better performance compared to the flow chart generated in Fig. 10. The model generates VFD O/W, which corresponds to an intermediate pattern delimited by TF and

D W/O, demonstrating correct behavior of the neural network. Moreover, it shows that only two patterns are generated outside the range delimited by the reference data, which are the VFD O/W and VFD W/O patterns. Furthermore, in the velocity range of 0.01 to 0.1 m/s for J_w and 0.09 to 0.1 m/s for J_o, the flow pattern S O/W is presented, for which the predictions developed by the implemented network show high precision when the injection velocities of water and oil are low, exhibiting only one error out of 52 predictions in this area. From this analysis, it is concluded that an increase in the oil injection velocity results in a displacement of the zone towards the vicinity of the transition zone. As a result, a greater imprecision is observed in the flow patterns obtained when using the LSTM 1 model.

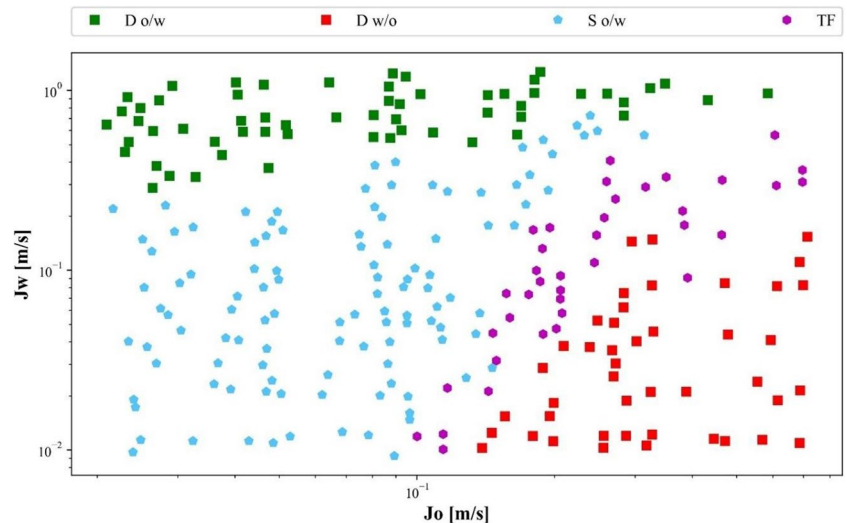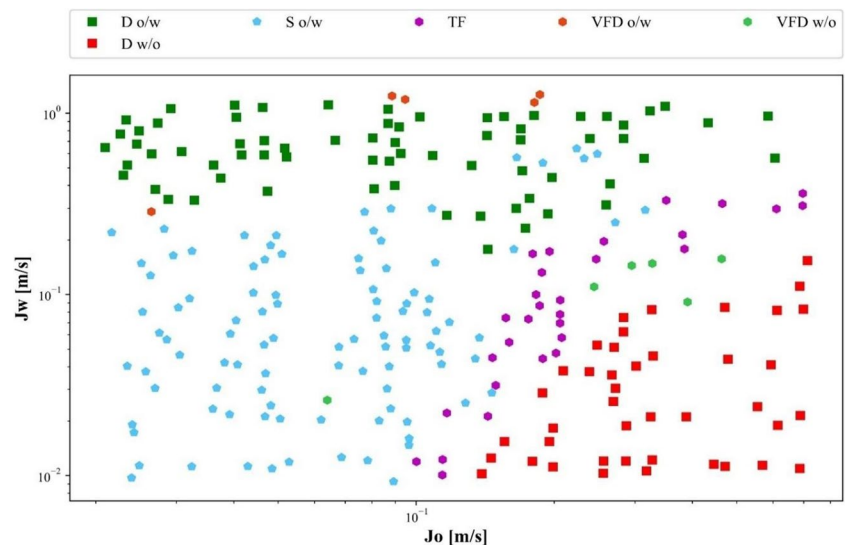**Fig. 11** Flow map with data from Mydlarz-Gabryk et al. [33]

Fig. 12 Flow map generated from Mydlarz-Gabryk et al. [33], applying the LSTM 1 model

4 Conclusions

In this study, various LSTM models were built for the prediction of data series containing liquid–liquid biphasic fluid characteristics of oil and water. These characteristics include oil and water injection rates, oil and water volume fractions, oil viscosity, and internal diameter of the pipe through which they flow. The objective was to determine the flow pattern present in a determined value of each characteristic. Four different configurations for the LSTM network were proposed, and once compiled and trained, the error was evaluated against the expected predictions. The results showed that LSTM 1 had the smallest error, with 0.98%, followed by LSTM 2 with 1.2%, LSTM 3 with 1.4%, and LSTM 4 with 3.5%. The LSTM 1 setup consisted of two hidden layers with 80 memory cells in

each, using categorical cross-entropy as the loss function. This allowed flow pattern prediction accuracy in a vertical pipe with less than 1% error.

The low performance of the LSTM 1 model in the transition zone is due to the scarcity of data available in these areas. In addition, it was shown that the performance of the model improves significantly when a large amount of data is available, increasing the accuracy from 50 to 83%. This highlights the importance of providing a larger amount of data for network training.

This study has shown that it is possible to develop models based on the prediction of time series using injection velocities, volume fractions, diameter, and viscosity, to determine liquid–liquid biphasic flow patterns of oil and water in a vertical pipeline. LSTM networks are particularly well-suited for this task due to their ability to model sequential

data, including time-dependent behavior in flow patterns. By structuring the data into temporal sequences, LSTM enables the discernment of patterns that other techniques might overlook, offering a detailed and insightful analysis of flow behavior. This represents a significant advantage in addressing the dynamic nature of flow patterns in complex systems. In the future, it is planned to investigate the behavior of the model with data from offshore platforms and use physically informed networks to obtain samples, with the aim of improving the quality and representativeness of the data used in network training.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. F. Amir, Y. Zhang, in *Perspectives on statistical thermodynamics*. Introduction to transport phenomena (Cambridge University Press, 2017), pp. 81–93. https://doi.org/10.1017/9781316650394.010

2. A. Archibong-Eso, N.E. Okeke, Y. Baba, A.M. Aliyu, L. Lao, H. Yeung, "Estimating slug liquid holdup in high viscosity oil-gas two-phase flow", Flow Meas. Instrum. **65**, 22–32 (2018). https://doi.org/10.1016/j.flowmeasinst.2018.10.027. (**Mar. 2019**)

3. A.C. Bannwart, O.M.H. Rodriguez, C.H.M. de Carvalho, I.S. Wang, R.M.O. Vara, Flow patterns in heavy crude oil-water flow. J. Energy Resour. Technol. **126**(3), 184–189 (2004). https://doi.org/10.1115/1.1789520

4. C.M. Ruiz-Diaz, M.M. Hernández-Cely, O.A. González-Estrada, Analysis of liquid-liquid (water and oil) two-phase flow in vertical pipes, applying artificial intelligence techniques. J. Phys. Conf. Ser. **2046**(1), 012016 (2021). https://doi.org/10.1088/1742-6596/2046/1/012016

5. G. Mask, X. Wu, K. Ling, An improved model for gas-liquid flow pattern prediction based on machine learning. J. Pet. Sci. Eng. **183**, 106370 (2019). https://doi.org/10.1016/j.petrol.2019.106370

6. E.M. Fayyad, A.M. Abdullah, M.K. Hassan, A.M. Mohamed, G. Jarjoura, Z. Farhat, Recent advances in electroless-plated Ni-P and its composites for erosion and corrosion applications: a review. Emergent Mater. **1**(1–2), 3–24 (2018). https://doi.org/10.1007/s42247-018-0010-4

7. M. Parsi, K. Najmi, F. Najafifard, S. Hassani, B.S. McLaury, S.A. Shirazi, A comprehensive review of solid particle erosion modeling for oil and gas wells and pipelines applications. J. Nat. Gas Sci. Eng. **21**, 850–873 (2014). https://doi.org/10.1016/j.jngse.2014.10.001

8. A.R. Hasan, C.S. Kabir, A simplified model for oil/water flow in vertical and deviated wellbores. SPE Prod. Facil. **14**(1), 56–62 (1999). https://doi.org/10.2118/54131-PA

9. M. Hotvedt, B.A. Grimstad, L.S. Imsland, Passive learning to address nonstationarity in virtual flow metering applications. Expert Syst Appl **210**, 118382 (2022). https://doi.org/10.1016/j.eswa.2022.118382

10. Y.A. Çengel, J.M. Cimbala, *Fluid mechanics: Fundamentals and applications*, 3rd edn. (McGraw Hill, New York, NY, 2014)

11. Q.B. Yong Bai, *Subsea Engineering Handbook*, 1st edn. (Elsevier, Houston, 2010). https://doi.org/10.1016/B978-1-85617-689-7.10028-7

12. X. Chen, L. Guo, Flow patterns and pressure drop in oil–air–water three-phase flow through helically coiled tubes. Int. J. Multiph. Flow **25**(6–7), 1053–1072 (1999). https://doi.org/10.1016/S0301-9322(99)00065-8

13. A. Faghri, Y. Zhang, *Fundamentals of Multiphase Heat Transfer and Flow* (Springer International Publishing, Cham, 2020)

14. A. Fallis, Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications **53**, 9 (2013)

15. C.M. Ruiz-Diaz, M.M. Hernández-Cely, O.A. González-Estrada, Modelo predictivo para la identificación de la fracción volumétrica en flujo bifásico. Cienc en Desarro **12**(2), 49–55 (2021). https://doi.org/10.19053/01217488.v12.n2.2021.13417

16. G. M. P. Andrade et al. Virtual flow metering of production flow rates of individual wells in oil and gas platforms through data reconciliation," *J. Pet. Sci. Eng* 208. 2022, https://doi.org/10.1016/j.petrol.2021.109772.

17. C.C. Aggarwal, *Neural Networks and Deep Learning*, vol. 1 (Springer International Publishing, Cham, 2018)

18. J. Sun, X. Ma, and M. Kazi, "Comparison of decline curve analysis DCA with recursive neural networks RNN for production forecast of multiple wells," *SPE West. Reg. Meet. Proc* 2018. https://doi.org/10.2118/190104-ms.

19. S. Song et al., An intelligent data-driven model for virtual flow meters in oil and gas development. Chem. Eng. Res. Des. **186**, 398–406 (2022). https://doi.org/10.1016/j.cherd.2022.08.016

20. N. Andrianov, A machine learning approach for virtual flow metering and forecasting. IFAC-PapersOnLine **51**(8), 191–196 (2018). https://doi.org/10.1016/j.ifacol.2018.06.376

21. C. M. Ruiz-Diaz, "Caracterización hidrodinámica de flujos multifase empleando técnicas de inteligencia artificial," Master Thesis, Universidad Industrial de Santander, Bucaramanga, 2021.

22. P. Abduvayt, R. Manabe, T. Watanabe, and N. Arihara, "Analysis of oil-water flow tests in horizontal, hilly-terrain, and vertical pipes," in Proceedings of SPE Annual Technical Conference and Exhibition, 2004. 1335–1347. https://doi.org/10.2523/90096-MS.

23. M. Du, N.-D. Jin, Z.-K. Gao, Z.-Y. Wang, L.-S. Zhai, Flow pattern and water holdup measurements of vertical upward oil–water two-phase flow in small diameter pipes. Int. J. Multiph. Flow **41**, 91–105 (2012). https://doi.org/10.1016/j.ijmultiphaseflow.2012.01.007

24. J. Flores, X. Chen, J. Brill, Characterization of oil-water flow patterns in vertical and deviated wells. Proceedings of SPE Annual Technical Conference and Exhibition **1**, 55–64 (1997). https://doi.org/10.2523/38810-MS

25. T. Ganat, S. Ridha, M. Hairir, J. Arisa, R. Gholami, Experimental investigation of high-viscosity oil–water flow in vertical pipes: flow patterns and pressure gradient. J. Pet. Explor. Prod. Technol. **9**(4), 2911–2918 (2019). https://doi.org/10.1007/s13202-019-0677-y

26. G.W. Govier, G.A. Sullivan, R.K. Wood, The upward vertical flow of oil-water mixtures. Can. J. Chem. Eng. **39**(2), 67–75 (1961). https://doi.org/10.1002/cjce.5450390204

27. Y.F. Han, N.D. Jin, L.S. Zhai, H.X. Zhang, Y.Y. Ren, Flow pattern and holdup phenomena of low velocity oil-water flows in a vertical upward small diameter pipe. J. Pet. Sci. Eng. **159**(May), 387–408 (2017). https://doi.org/10.1016/j.petrol.2017.09.052

28. A. R. Hasan and C. S. Kabir, "New model for two-phase oil/water flow: production log interpretation and tubular calculations," *Soc. Pet. Eng. AIME, SPE*, PI, 369–382 18216 1988, https://doi.org/10.2118/18216-pa.

29. A.K. Jana, G. Das, P.K. Das, Flow regime identification of two-phase liquid–liquid upflow through vertical pipe. Chem. Eng. Sci. **61**(5), 1500–1515 (2006). https://doi.org/10.1016/j.ces.2005.09.001

30. A.K. Jana, P. Ghoshal, G. Das, P.K. Das, An analysis of pressure drop and holdup for liquid-liquid upflow through vertical pipes. Chem. Eng. Technol. **30**(7), 920–925 (2007). https://doi.org/10.1002/ceat.200700033

31. J. Guo et al., Heavy oil-water flow patterns in a small diameter vertical pipe under high temperature/pressure conditions. J. Pet. Sci. Eng. **171**, 1350–1365 (2018). https://doi.org/10.1016/j.petrol.2018.08.021

32. R.A. Mazza, F.K. Suguimoto, Experimental investigations of kerosene-water two-phase flow in vertical pipe. J. Pet. Sci. Eng. **184**, 106580 (2020). https://doi.org/10.1016/j.petrol.2019.106580

33. K. Mydlarz-Gabryk, M. Pietrzak, L. Troniewski, Study on oil-water two-phase upflow in vertical pipes. J. Pet. Sci. Eng. **117**, 28–36 (2014). https://doi.org/10.1016/j.petrol.2014.03.007

34. O.M.H. Rodriguez, A.C. Bannwart, Experimental study on interfacial waves in vertical core flow. J. Pet. Sci. Eng. **54**(3–4), 140–148 (2006). https://doi.org/10.1016/j.petrol.2006.07.007

35. J. Xu, D. Li, J. Guo, Y. Wu, Investigations of phase inversion and frictional pressure gradients in upward and downward oil–water flow in vertical pipes. Int. J. Multiph. Flow **36**(11–12), 930–939 (2010). https://doi.org/10.1016/j.ijmultiphaseflow.2010.08.007

36. Y. Yang et al., Oil-water flow patterns, holdups and frictional pressure gradients in a vertical pipe under high temperature/pressure conditions. Exp. Therm. Fluid Sci. **100**, 271–291 (2018). https://doi.org/10.1016/j.expthermflusci.2018.09.013. (**Jan. 2019**)

37. D. Zhao, L. Guo, X. Hu, X. Zhang, X. Wang, Experimental study on local characteristics of oil-water dispersed flow in a vertical pipe. Int. J. Multiph. Flow **32**(10–11), 1254–1268 (2006). https://doi.org/10.1016/j.ijmultiphaseflow.2006.06.004

38. J. Brownlee, *Long short-term memory networks with python develop sequence prediction models with deep learning* (Machine Learning Mastery, Vermont, 2017)

39. K.K. Chandriah, R.V. Naraganahalli, RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting. Multimed. Tools Appl. **80**(17), 26145–26159 (2021). https://doi.org/10.1007/s11042-021-10913-0

40. C. Ruiz-Díaz, M.M. Hernández-Cely, O.A. González-Estrada, Modelo predictivo para el cálculo de la fracción volumétrica de un flujo bifásico agua- aceite en la horizontal utilizando una red neuronal artificial. Rev UIS Ing **21**(2), 155–164 (2022). https://doi.org/10.18273/revuin.v21n2-2022013