The application of deep learning algorithms to classify subsurface drilling lost circulation severity in large oil field datasets

Sajjad Mardanirad¹ · David A. Wood² · Hassan Zakeri³

Received: 15 June 2021 / Accepted: 11 August 2021

Published online: 26 August 2021

© The Author(s) 2021

Abstract

In this paper, we present how precise deep learning algorithms can distinguish loss circulation severities in oil drilling operations. Lost circulation is one of the costliest downhole problem encountered during oil and gas well construction. Applying artificial intelligence can help drilling teams to be forewarned of pending lost circulation events and thereby mitigate their consequences. Data-driven methods are traditionally employed for fluid loss complexity quantification but are not able to achieve reliable predictions for field cases with large quantities of data. This paper attempts to investigate the performance of deep learning (DL) approach in classification the types of fluid loss from a very large field dataset. Three DL classification models are evaluated: Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU) and Long-Short Term Memory (LSTM). Five fluid-loss classes are considered: No Loss, Seepage, Partial, Severe, and Complete Loss. 20 wells drilled into the giant Azadegan oil field (Iran) provide 65,376 data records are used to predict the fluid loss classes. The results obtained, based on multiple statistical performance measures, identify the CNN model as achieving superior performance (98% accuracy) compared to the LSTM and GRU models (94% accuracy). Confusion matrices provide further insight to the prediction accuracies achieved. The three DL models evaluated were all able to classify different types of lost circulation events with reasonable prediction accuracy. Future work is required to evaluate the performance of the DL approach proposed with additional large datasets. The proposed method helps drilling teams deal with lost circulation events efficiently.

Article Highlights

- Three deep learning models classify fluid loss severity in an oil field carbonate reservoir.
- Deep learning algorithms advance machine learning a large resource dataset with 65,376 data records.
- Convolution neural network outperformed other deep learning methods.

Keywords Lost circulation · Machine learning · Deep learning of drilling datasets · Convolutional neural network · Gated recurrent unit · Long-short term memory

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42452-021-04769-0>.

✉ Sajjad Mardanirad, mardanirad1@gmail.com; David A. Wood, dw@Dwasolutions.com; Hassan Zakeri, hzakeri@pasargadep.com | ¹Oil Industries Engineering and Construction Group, Tehran, Iran. ²DWA Energy Limited, Lincoln, United Kingdom. ³Pasargad Exploration and Production, Tehran, Iran.

1 Introduction

Drilling oil and gas wells is associated with a variety of downhole problems. The cost of lost drilling fluid and treatment of circulating system is a burden for the drilling industry. Proper drilling-fluid treatment planning and swift response of rig crew can prevent subsurface fluid losses and diminish further problems. Downhole losses can trigger complications such as well control and stuck pipe challenges which are time and cost consuming issues. Also, utilizing improper treatment solutions can lead to wellbore instability and formation damage by plugging hydrocarbon fluid conduits which cause production constraints [1]. Losing drilling mud through subsurface formations, which is called lost circulation, occurs with different severities. The loss severity is categorized according to loss rates varying from seepage, partial loss, severe loss, and ultimately reaching complete loss of drilling fluid.

Loss circulation occurs downhole usually associated with two distinct conditions: 1) the wellbore pressure exceeds the pore pressure of the subsurface formations, or, 2) a large subsurface void space (pore or fracture related) is encountered that consumes wellbore fluids [2, 3]. Lost circulation is derived from natural and induced fractures, cavernous or vugular formation textures or a mixture of these characteristics, as depicted in Fig. 1. Fluid loss severity could be minor, during which drilling is able proceed, or severe to complete, where the primary objective becomes focused on safeguarding / rescuing the well. During overbalanced subsurface pressure conditions in a well, the natural fractures, cavernous or vugular void spaces, and/or poorly consolidated lithologies take drilling fluid from the wellbore, particularly while the drilling bit is in contact with such formations. Sometimes, the void spaces are extensively distributed throughout the formations being drilled such that that even underbalanced systems can lead to mud losses. This typically occurs because natural fractures do not have sufficient strength to bear the well fluid pressure resulting in tensile failure. One approach to dealing with lost circulation is to continue drilling in blind mode. This is commonly the response in offshore wells, where drilling is continued until the hole section reaches its planned depth. At that point, the lost zone is plugged. Severe to complete fluid losses conventionally are restrained by setting a cement plug or gunk pill across thief zone. A general cause of induced mud losses is inappropriate mud weight selection, although induced failures can arise due to unexpected formation anomalies. It results in the generation of induced fractures because the formation strength is insufficient to

Fig. 1 Different subsurface systems encounter loss circulation of drilling fluid

bear fluid pressure leading to drilling fluid (mud) losses to the formation [2, 4].

As a general rule, a range of lost circulation materials (LCM) are added to the drilling fluid to fill the void spaces, in attempts to cure or prevent lost circulation occurring in the vugs and fractures that characterize some rock formations. However, the traditional method of selecting plugging agents is not effective in many cases, because the LCM types are selected by trial and error [5]. Implementing statistical methods using actual operational data can help to mitigate potential risks of drilling fluid losses from occurring. Such methods attempt to predict the subsurface zones particularly susceptible to lost circulation and classify them in terms of the loss severity most likely to occur.

In this study, we apply three deep learning methods to a large drilling dataset from the giant Azadegan oil field onshore Iran. Section 2 highlights the significance, impacts and financial burden of lost circulation. Section 3 describes in outline the methods evaluated and the dataset studied. Section 4 reviews existing lost circulation models. Section 5, supported by Appendix 1, provides the theoretical background to the deep learning methods evaluated. Section 6 describes in detail the variables considered in the dataset. Section 7 explains how the data records are pre-processed and the deep learning models are constructed

and their prediction performance evaluated. Section 8 presents and discusses the results of applying the deep learning methods developed to the Azadegan dataset.

2 The impacts of lost circulation on drilling operations

Reducing the required budget per well for all oil field developments is a priority. Lost circulation is a major challenge during well construction that annually results in considerable and costly nonproductive rig time. The global cost associated with lost circulation is estimated to be hundreds of million dollars, which includes lost drilling fluids and cost of treatments, as well as the lost drilling time and cost of specialist tools and materials involved in its remediation. Since, nearly 10% of the well costs are associated with drilling mud cost, downhole losses significantly escalate well costs [6, 7]. For example, the total cost imposed on drilling operators due to lost circulation in the wells drilled in the Azadegan oilfield, considered here, is estimated to be about 5.35 million dollars. This issue results in considerable non-productive time (NPT) and other downhole problems. The NPT associated with different problems caused by lost circulation, expressed in percentage terms, is categorized in Fig. 2.

3 Method and materials

Implementing robust algorithms with the capability of efficiently processing large dataset (i.e., with tens of thousands of data records) to classify various classes of downhole lost circulation remains a priority for large oil fields suffering significant lost circulation issues in specific reservoir zones. Deep learning (DL) algorithms offer the capability to do this but have yet to be widely applied

for that purpose. This study achieves this objective with a large, compiled data set for the giant Azadegan oil field (onshore Iran). This oil field is yet to be fully developed, and requires many more future development wells to be drilled. Hence, the method proposed should benefit that field by providing improved lost circulation predictions.

Supervised deep and machine learning (DL/ML) solutions for many downhole problems, such as lost circulation, are challenging because of subsurface environmental complexity. In such situations data extraction plays an important role in DL/ML model success. DL architectures assist in data selection by implementing high-level analysis of a dataset taking into account the most influential features. Here, the features of the dataset were specifically selected based on field experience. The features considered include both drilling parameters and mud rheology conditions in an effort to identify a general classifier for lost circulation problems with large dataset. The robustness of DL tends to be better than for conventional ML algorithms when the task involves solving problems with large data amount.

Most of previous works predicted and classified lost circulation with various ML algorithms based on limited datasets. Although, the ML algorithms applied achieve good predictions for the small datasets evaluated they tend not to perform very well when applied in different locations. So, finding a powerful and unique solution to this matter remains an important goal for the industry. Here, three DL methods are used to classify different types of drilling fluid lost circulation events: Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU) and Long-Short Term Memory (LSTM). These models are constructed using the TensorFlow programming library [8]. The DL models were all trained and tested with data records from the Azadegan oil field. The performance of the models in classifying types of fluid losses was evaluated using the commonly applied accuracy, precision, recall and F1-score metrics.

Fig. 2 Non-productive time (NPT), the consequences of lost circulation problems in the Azadegan oilfield

Figure 3 illustrates the lost circulation classification workflow applied in this study. The DL approach to mud losses classification proposed and evaluated should be analyzed further with different dataset to come up with a general and unique solution.

The recurrent neural network (RNN) architectures are commonly employed with time series data. However, the downhole data applicable to lost-circulation analysis consists of depth-based records. In this study such data records are gathered from several wells to compile a single dataset. The powerful features of RNN are exploited to overcome the weakness of traditional ML methods in classification of large datasets.

Firstly, the required data records are extracted from an extensive subsurface database which include real time parameters of all the wells drilled to date in Azadegan oil field. Preprocessing removes any bad records and the viable real time parameters are compiled. Then, precise mud properties are added to the dataset. This involves a comprehensive investigation of daily drilling and mud reports and a review of each final well report. Each data record is labelled with a classification relating to fluid loss conditions using information extracted from the daily drilling and final well reports.

The data records are normalized as part of the pre-processing analyses to avoid biases caused by the different scales of individual variables. During this step,

no dimension reduction (preferential feature selection) is applied to available recorded variables, since the DL methods considered involve routines that enable them to identify and ignore irrelevant variables, i.e., those found not to be useful for the desired classification task. An early-stopping approach was established to minimize the risk of overfitting by the ML models. The deep learning networks are developed by splitting the normalized data records into training (~ 75%) and testing (~ 25%) subsets. In traditional feed-forward architectures each input data value acts independently, whereas in time-series networks each input value is dependent on the previous input for that variable. Therefore, the models are constructed to feed input data values, item by item in a sequential series. They are firstly trained to distinguish five defined classes of lost circulation events and their classification performance is subsequently verified using the independent testing subset. In order to compare the performance of LSTM, GRU, and CNN models, the number of epochs (iterations) executed was set equal for all models.

Fig. 3 Lost circulation classification workflow

4 Lost circulation models

4.1 Analytical models

Selecting the optimum treatment method requires the knowledge of the severity of downhole drilling fluid losses. Several attempts have been made to accurately predict the amount of drilling fluid losses into subsurface formations. Parn-anurak and Engler [9] applied Darcy's law to propose a numerical relationship for estimating the volume of drilling fluid that infiltrates into the permeable portions (void spaces) of rocks. Formations with natural fractures tend to be particularly prone to lost circulation with various degrees of severities. Numerous studies assess and attempt to estimate the degree of the lost circulation volumes into fractured media using quantitative approaches [10–18]. These studies have led to a simplified formula being developed to describe the lost volume in terms of the Yield Power Law (or Heschel Buckley formula) applied to drilling fluids in natural fractured formations, assuming a constant fracture width and ignoring the ballooning effect [19]. Since the total loss circulation has a complex mechanism and usually occurs in formations with complex structures such as caverns or large fractured networks, particularly carbonates, no quantitative formulation has yet been developed to describe this phenomenon.

4.2 Machine learning models

Prediction of downhole problems requires offset well data and sufficient data records to provide classification with statistical confidence. In the past decade, ML has been implemented as a key tool in the oil and gas industry. Several authors have studied the performance of machine learning algorithms for reservoir simulation, history matching, and production optimization. The performance of this approach in different aspects of the drilling industry has also been evaluated by numerous researchers. Mohaghegh [20] developed an integrated system that supports the drilling operations with the aim of reduction in non-productive time and enhancement in drilling performance using artificial neural networks (ANNs). Hedge and Grey [21, 22], Dunlop et al. [23], Yavari et al. [24], and Sabah et al. [25] have established machine learning models with real-time data records to optimize drilling rate of penetration (ROP). Castiniera et al. [26] provided a data-driven workflow to enhance drilling performance using a natural language processing approach. Kowalchuk [27] also demonstrated the advantages of natural language processing techniques to extract information from drilling reports.

The support vector machine (SVM) algorithm has also been applied to geological data in attempts to optimize the well trajectory to a reservoir target in the exploration wells [28]. With a similar objective, Pollock et al. [29], implemented a reinforcement learning algorithm to attain the optimum directional well path with minimum deviation from the planned trajectory. Zhang et al. [30] conducted risk analysis using machine learning algorithms associated with managed pressure drilling (MPD) systems. Whereas, Sule et al. [31] performed quantitative risk analysis to assess the kick control capabilities of an MPD system using a dynamic Bayesian network. Bhandari et al. [32] reported that a Bayesian network could also usefully assist with well blow-out predictions in offshore fields. Hoffiman et al. [33] applied various deep learning algorithms to data extracted from drilling reports to characterize different operational sections of wellbores. Based on their results, the LSTM network showed a more reliable performance for that purpose compared to ANN and CNN algorithms.

Alhameedi et al. [34, 35] exploited regression analysis to provide a statistical formula involving various drilling parameters known to exhibit influential relationships with lost circulation. However, the application of various machine learning modeling methods for predicting downhole lost circulation with improved accuracy has been the focus of several recent studies. Toreifi et al. [36] identified the advantages of the SVM algorithm in comparison with multi-layer perceptron (MLP) for prediction of loss events in fractured formations. Cristofaro et al. [37] implemented several machine learning algorithms to classify the success in applying different LCMs in a deep-water pre-salt basin offshore Brazil. They showed that by combining neural networks with an instance-based algorithm better prediction performance could be achieved. Two studies have identified the good prediction performance of SVM of lost circulation events during drilling in underbalanced mode [38, 39].

A smart loss circulation volume estimation method has been developed [40] using adaptive neuro-fuzzy inference systems (ANFIS), coupled with genetic-multilayer perceptron (GA-MLP), artificial neural networks and decision tree (DT) algorithms. Based on the obtained results, the DT algorithm achieved the best performance accuracy in that study. A lost circulation prediction approach based on seismic data anomalies using machine learning was developed by Geng et al. [41]. Alkinani et al. [42] established a data-driven neural network using the Bayesian Regularization (BR) algorithm to control the equivalent circulation density (ECD) variable in order to minimize non-productive time and cost. Li et al. [43] performed risk analysis on lost circulation during drilling utilizing various supervised machine learning algorithms. The best

performance attained in that study was associated with the Random Forest (RF) algorithm.

Ahmed et al. [44] evaluated ANN and SVM algorithms for intelligent prediction of lost circulation and observed that the performance of the SVM algorithm was superior to their ANN model. Alkinani et al. [2] demonstrated that the Levenberg-Marquardt (LM) algorithm performs better than other ANN training algorithms in predicting the lost circulation through induced fractures. Hou et al. [45] also developed an ANN model with satisfactory outcomes for predicting lost circulation in an offshore high-pressure high-temperature (HPHT) field. However, the ANN and SVR approaches typically do not perform well for lost circulation in natural and hydraulically induced fractures associated with large datasets [46]. Alkinani et al. [1] utilized different data driven algorithms applied to a large dataset of 3000 wells where a Quadratic SVM model achieved the best prediction performance. The BiGAN & 1D CNN model was coupled by Wang et al. [47] using time series data to predict loss circulation. In that study, the models were fed with the real-time data of a 16.7-min drilling period (recorded every 0.02 s) with high accuracy. Ljubran et al. [48] also investigated the role of Random Forest (RF), Artificial Neural Network (ANN) and DL including LSTM and CNN models to predict lost circulation. That work revealed the superiority of CNN over other algorithms. Sabah et al. [49] utilized the hybrid ML algorithms where the combinations of least square support vector machine and cuckoo search algorithm (LSSVM-COA) and least square support vector machine and particle swarm size (LSSVM-PSO) outperformed other algorithms. The DL/ML published studies on lost circulation prediction and classification are summarized in Table 1.

5 Deep learning theory

Deep learning (DL) has achieved more attention in recent years because of improved performance and accessibility of such algorithms. The robustness of DL algorithms over conventional ML networks has been identified in many studies. Osarogiagbon et al. [56] illustrated the performance curve of DL algorithms in comparison to conventional ML and neural networks according to the amount of data processed. Figure 4 shows schematically that for studies involving large datasets, DL algorithms tend to outperform ML algorithms.

In 2017, Google's brain team released the first version of its deep learning framework in providing access to its more advanced learning algorithms [57]. In particular, the TensorFlow engine is now established as a library [8] which enables engineers to apply complex network architectures, regardless of the amount of data records to be

processed. In this study, the TensorFlow library is used to provide the deep learning engines applied to solve the lost circulation classification challenge with large well datasets. The analysis also meaningfully compares the classification performance and processing speed of two deep learning networks, namely RNN and CNN.

5.1 Long short-term memory (LSTM)

History data play a significant role in forecasting a time-series process, such as for the prediction of stock market indices or making power consumption forecasts in specific regions based on underlying weather and environmental variables [58]. The long short-term memory architecture allows the past results to flow through the deep learning network structure and continuously contribute to each prevailing iteration. This mechanism makes the model flexible, thereby contributing to the convergence of the model and maintaining its dependency on the distant past data processed in much earlier iterations [59]. Figure 5 illustrates how an LSTM cell works with more details of the method provided in Appendix 1.

5.2 Gated recurrent unit (GRU)

Since the LSTM cell has a complicated structure, a simplified version of the LSTM network has been introduced that is termed the Gated Recurrent Unit (GRU) [60]. Two gates are used to construct the architecture of a GRU cell. The data that flows into and out of a GRU memory cell are controlled by an update, z_t , and a reset gate, r_t , respectively. Figure 6 illustrates the structure of a GRU cell. In the research carried out by Gao and Dorota [61], LSTM and GRU exhibited almost identical performances in terms of the prediction accuracy they achieved. However, the GRU algorithm computes faster than LSTM, due to its simplified structure. More details of the method provided in Appendix 1.

5.3 Convolutional neural network (CNN)

Convolutional Neural Networks (CNNs) were first introduced in the 1980s [59]. They are implemented as a powerful tool for training unstructured data using the convolution concept. This method enables engineers to solve a variety of perceptual tasks and is popular for its ability to detect objects within picture, video or image segments. CNN algorithms are now applied in many routine daily life applications [62], because CNNs typically yield higher quality results than other deep learning networks. One advantage of CNNs is that data are localized step by step. This means that CNN training can be performed using the minimum amount of data, resulting in less computational

Table 1 Lost circulation predictions using different artificial intelligent methods

Algorithms artificial intelligence	Input parameters	Authors	Data availability	Performance
ANN	Depth, formation top, coordination, bit size, pump output, pump pressure, MW, solid percent, and fluid loss of previous time frame	Moazzeni et al. [50]	Data of 32 wells used as input parameters Data splitting as: Training: 70% Validation: 15% Testing: 15%	Training R value: 0.95 Validation R value: 0.82 Testing R value: 0.76
ANN	Coordinates, the current depth, formation top, ROP, formation type, annulus volume, mud pressure, flow rate, pump pressure, filter cake viscosity, solid content, PV, YP, initial GS and 10 min GS	Toreifi and Rostami, [36]	1630 data records from 38 drilled wells Data splitting as: Training: 60% Validation: 20% Testing: 20%	Modular neural network R value: 98%
Bayes discriminant	Casing pressure and SPP	Liang et al. [51]	Training data interval: 3123.5 to 3133.6 m Testing data interval: 3122.1 to 3131.4 m The prior probabilities have attained from literature/expert view	The kick and loss detected with change in casing pressure and SPP using the Bayes discriminant algorithm The loss circulating risk evaluated for the states: "not circulating", "tripping in" and circulating while the circulating scenario acquired the highest occurrence probability of lost circulation
Dynamic Bayesian network	MW, flow meter failure, natural micro fracture, low formation fracture pressure, formation porosities, casing failure, cement failure, running drill pipe rate, temperature, drilling fluid viscosity, pump out, high pump pressure	Wu et al. [52]		
ANN	MW, depth, pump pressure, pump flow rate	Behnoudfar and Hosseini, [38]	Data of 3 wells used in this study Data splitting as: Training: 75% Validation: 15% Testing: 15%	Observed that pump pressure increases results in lost circulation rise up, indeed, MW decreases by increase in pump flow rate decreases the MW
Naive Bayes, Instance Based, and ANN	The static and dynamic lost circulation rate, coordinates, the current well depth, the preferred lost circulation material	Cristofaro et al. [37]	The third-party software database was used in this study	The best result attained in the combination of the MLP Neural Network with KStar instance-based classifier with ROC: 94.1%
ANFIS	Drilling footage, drilling time, hole size, WOB, RPM, pump rate, pump pressure, mud viscosity, shear stress (shear rate 300&600), GS, solid content, pore pressure, mud pressure, fracture pressure and fluid lost amount	Agin et al. [53]	42,948 data of 61 wells of Maroon oil field were used to predict the lost circulation Data splitting as: Training: 66% Validation: 17% Testing: 17%	Training RMSE: 0.083 Validation RMSE: 0.154 Testing RMSE: 0.091
Logistic Regression, Random Forest, C-Support Vector which adopted majority voting algorithm	Depth, time, variance, attenuation, sweetness, RMS amplitude	Geng et al. [41]	Not available	The performance evaluated by Cross-validation score: 0.8 (± 0.0021)
Bayesian inference algorithm	Mud flow-out rate, flow-in rate, SPP, total pit volume	Wu et al. [54]	Time-based data of 5 s which time ranged from 20:20:07 to 22:34:57 as validation datapoints	Concluded that combining dynamic threshold-based diagnosis and Bayesian estimation can improve model performance

Table 1 (continued)

Algorithms artificial intelligence	Input parameters	Authors	Data availability	Performance
ANFIS, DT, MLPNN, RBF-NN and GA-MLP	Depth, coordination, hole size, drilling time, drilling meterage, WOB, SPP, RPM, flow-in, viscosity, shear stresses (300&600 RPM), GS, solid percent, formation type, pore pressure, drilling mud pressure and formation fracture pressure	Sabah et al. [40]	Data from 61 wells including 1900 datapoints Data splitting as: Training data: 70% Testing data: 30%	RMSE/R metric values: DT Test: 0.091 / 0.93 DT Train: 0.052/0.9725 MLP-NN Test: 0.099/0.9073 MLP-NN Train: 0.094/0.9213 ANFIS Test: 0.1087/0.8855 ANFIS Train: 0.1163/0.9018 RBF-NN Test: 0.1315/0.8445 RBF-NN Train: 0.1172/0.85 GA-MLP Test: 0.137/0.826 GA-MLP Train: 0.132/0.8307 SVM model training R2 value: 0.97 and testing R2 value: 0.95
ANN and SVM	Measured depth, hole size, inclination, azimuth, WOB, RPM, MW, Flow-in, ROP, circulating pressure, lithology, solid content, fluid loss, YP, PV, funnel viscosity, 10-s. GS, 10-min. GS	Abbas et al. [55]	1120 datapoint used in this study Data splitting ratio of 3:1 for training and testing	
SVM, decision trees, logistic regression, ANNs, and ensemble trees	Reason of loss (vugs and caves, natural fractures, induced fractures), type of treatment, type of loss, well type	Alkinani et al. [1]	Utilized data for more than 3000 wells	Quadratic SVM had highest performance with 74% in accuracy metric
BiGAN & 1D CNN	SPP, annulus pressure, WOB, TOB, annulus temperature, X-axis vibration (XV), Y-axis vibration (YV) and Z-axis vibration (ZV)	Wang et al. [47]	The real-time data of a 16.7-min drilling period, recording every 0.02 s	Mud loss in normal condition obtained 99% accuracy. According to the probability line, loss event was commenced at 199,225 s (5.9 min earlier than the reported detect)
Random forest (RF), ANN, CNN and LSTM	WOB, hook height, hook load, torque, SPP, flow-in, flow-out, ROP, RPM, PVT, trip tank volume	Ljubran et al. [48]	The 48,000 data records were gathered for loss circulation detection from ~200 wells with an average frequency of 0.2 Hz/record. For each well 60 and 180 samples are extracted for LCIs and normal operations, respectively. Data splitting as: Training: 80% Validation: 10% Testing: 10%	CNN models performed best in the early detection of LCIs with precision, recall, and F1-scores of 87.34%, 73.40%, and 79.77%, respectively
Genetic algorithm (GA), particle swarm size (PSO), cuckoo search algorithm (COA), multilayer perceptron (MLP) neural network and least square support vector machine (LSSVM)	Easting, northing, depth, drilling meterage, drilling time, formation type, hole size, WOB, flow rate, mud weight, bit rotational speed, pump rate, pump pressure, formation and mud properties	Sabah et al. [49]	2820 cases of 305 drilled wells in the Marun oil field implemented	LSSVM-COA and LSSVM-PSO was two most accurate models with coefficients of determination of 0.9424 and 0.9391, respectively

time [63]. The generic CNN architecture is illustrated in Fig. 7.

Typically, a set of connected network section including input, convolution, pooling, flatten, and logit layers constitute the architecture of a CNN model.

During the convolution phase, certain data variables according to specified probabilities are progressively withdrawn from the network through a dropout layer, thereby enhancing the effectiveness of the learning process and reducing the risk of overfitting. The CNN only retains data variables essential for achieving accurate predictions of the dependent variable and also removes the noisy (outlying) data records. In order to perform the optimization of deep learning algorithms effectively, an appropriate loss function should be implemented according to the problem in hand. Since a multiclass dataset used in this paper, the categorical cross-entropy loss function has been chosen for RNN and CNN models construction. More details of the method provided in Appendix 1.

6 Dataset description

A machine can learn most patterns with an acceptable precision if it is fed with an adequate number of clean data records. Employing robust algorithms with sufficient data to predict various drilling problems greatly assists engineers in well planning and in minimizing the risk of encountering downhole problems. In this study, a large subsurface dataset has been compiled from 20 wells drilled into the Azadegan oilfield, Iran, which is the world's largest unexploited oil field [64–66]. The stratigraphic column for the Azadegan oilfield is shown in Fig. 8. The Sarvak carbonate reservoir contains most oil which is heavy and with about 19.95 deg API gravity (0.9343 relative density and 0.13%mol H_2S content). Approximately 200 wells have been drilled in this field so far. The deepest well drilled in this field has a vertical depth of 4850 m. This field is divided into three distinct structural blocks, I, II, and III. The dataset has been gathered from wells distributed across the field. Most of lost circulation events recorded have occurred in block III. The dataset evaluated includes 65,376 records (rows) of depth-based (meter by meter) mechanical parameters and drilling fluid rheological data acquired through investigating daily drilling, mud logging and final well reports. This dataset is made accessible to readers (see Appendix 2) Table 2 presents the value ranges of the variables associated with the data records in the dataset. In this table, the rotation parameter is defined as the revolution of the drilling string per minute and contains both surface and downhole motor revolutions in the available cases. Moreover, the loss rates have been extracted from daily drilling reports for each meter of the drilled depth and then converted to five classes varying from

Fig. 4 A schematic diagram showing the relative performance curves versus available data for different ML algorithms (Osarogiabon et al. [56])

Fig. 5 Generic configuration of a long-short term memory (LSTM) cell (see text for definition of each symbol)

Fig. 6 Generic configuration of a gated recurrent unit (GRU) cell (see text for symbol definitions)

Fig. 7 Generic configuration of a convolutional neural network architecture

“no loss” to “complete loss”. The label for “no loss” is coded as the number 0, whereas the label “complete loss” is coded as 5. described in footnotes to Table 1. Some seepage loss occurs in most of the field formations. However, Pabdeh-Jahrum, Pabdeh, Gurpi, Sarvak, and Gadvan are associated with the partial loss circulation classification. Severe Losses have been recorded in the Jahrum, Pabdeh-Jahrum, and Sarvak formations. These are predominantly carbonate lithologies. The Jahrum formation is a dolomitic lithology that is commonly associated with complete losses in this field. Since most formations of this oilfield experience some lost circulation issues, implementing a smart and comprehensive lost circulation predictor has become a priority task for improving drilling performance. Thus, the deep learning approach has been used here to classify the loss severity, minimize the risk of loss circulation occurring during drilling operations, and allowing rapid and appropriate remedial actions to be taken whenever it is encountered.

7 Data preprocessing and model development

7.1 Missing or erroneous data

Real-time data quality control was performed to filter out the outlying and noisy data records. These abnormal data were primarily associated with Flow In, Flow Out, Standpipe Pressure, and Pump Stroke parameters. These recording anomalies occurred due to failure or problems with the pumping system or inappropriate installation of the data acquisition system. Consequently, 1240 relevant

rows relating to poor quality data were removed to refine the dataset.

7.2 Data formatting

DL/ML methods typically require data variables recorded in numerical formats (*not text*). Thus, in order to make valid and precise predictions with data mining and deep learning methods, it is important to convert text data variables into numerical formats. Some converters are available for this purpose, for instance the Pandas Python package that has been implemented here to transform the textual features into numeric formats. Such conversion is applied to the Label features.

7.3 Standard normalization

From Table 2, it is apparent that the dataset comprises several features expressed on different scales. Large datasets with multiple variables expressed in different scales / units can be problematic and difficult to handle with a risk of providing biased results [67, 68]. Thus, all data variables were normalized to a scale of -1 and $+1$ prior to being analyzed by the deep learning models. This was achieved applying Eq. (1):

$$X_{norm} = 2 \frac{X - X_{min}}{X_{max} - X_{min}} - 1 \quad (1)$$

Age	Fm./Mem ber	Formation Top					Apparent Thick.	Lithology	Lithology Description
		Forecast		Graphic well log		Log			
		TVD	MD	TVD	MD	MD			
Pliocene	Aghajari	10.5	10.5	10.5	10.5	NA	952.5		Sst: loose sand, vfn-fn gr
									Clst: Rd., clac., sft., occ. Frm., stky.
									Clst: Gy., calc., sft., occ. Frm., stky.
									Anhy: Offwht., - wht., md. hd.
									Clst: Gy. calc., sft., occ. frm., stky.
Miocene	Gachsaran	978	978	963	963	960.7	308		Mrl: Dk. gr.- gr., sft., hi. calc.
	Asmari	1294	1294	1271	1271	1296.0	231		Lst: offwht., Lt. br., cr.-br., packstone, mud stone, type I/II, argi.
Paleocene-Eocene	Pabdeh-Jahrum	1502	1502	1502	1502	1502.0	129		Lst: wht., wht.-lt. gr., gr., type I/II
	Jahrum	1631	1631	1631	1631	1593.0	230		Dolomite
	Pabdeh	1866	1866	1861	1861	1918.5	273.5		Lst: wht., wht.-lt. gr., gr., type I/II
Upper Cretaceous	U. Gurpi	2135	2135	2134.5	2134.5	2133.1	181.5		Lst: wht, gr., lt. gr., dk. gr., mudstone, wackestone
	Tarbur	2316	2316	2316	2316	2325.5	89		type I/II, II, argi., pyr., rexln.
	L. Gurpi	2406	2406	2405	2405	2396.5	134.7		Mrl: gr, stf, stk, calc
	Ilam	2540	2555	2540	2555.5	2559.5	66.7		Lst: offwht, ltbr-br, gr, type I/II, II
	Laffan	2608	2660	2606.4	2666	2665.2	7.9		Mrl & shale
Lower Cretaceous	S1	2617	2680	2614.3	2683	2679.0	21.3		Lst: offwht., lt.cr.-xln., Lt. gr.-gr.-dk. gr., type I/II, II, I/I
	S2	2636	2732	2635.6	2744	2740.3	7.7		Lst: offwht., lt.cr.-xln., Lt. gr.-gr.-dk. gr., type I/II, II, I/I
	S3	2644	2766	2643.3	2772	2757.2	33.2		Lst: offwht., lt.cr.-xln., Lt. gr.-gr.-dk. gr., type I/II, II, I/I

Fig. 8 Statigraphic column of the Azadegan oilfield

7.4 Heat mapping the correlation coefficient

Correlation coefficient, R , was computed between all data

features and displayed in the form of a heat map diagram (Fig. 9). The strongest positive correlations are expressed in yellow and orange colors, whereas strong negative

Table 2 Variables and their value ranges for the 65,376 data records compiled for the Azadegan oilfield (Iran)

Feature	Range
Hole section (inch)	26–4.125
Depth (m)	14–4285
Rate of penetration (m/hr)	< 1–616
Weight on bit (klbs)	0–79
Rotation (RPM)	0–457
Torque (klbs-ft)	0–12.3
Standpipe pressure (psi)	9–3915
Flow in (gpm)	< 10–1934
Flow out (%)	0–100
Pump stroke (spm)	6–242
Mud weight (pcf)	66–127
Funnel viscosity (seq/qt)	29–76
Plastic viscosity (cp)	3–48
Yield point (lbs/100 ft ²)	2–47
Gel strength 10 s (lbs/100 ft ²)	1–26
Gel strength 10 min (lbs/100 ft ²)	1–27
Solid (%)	4–65
Losses Severity ^a	[5 Classes]

Losses Severity^a No loss–Seepage–Partial–Severe–Complete

correlations are expressed in purple and blue colors. Most of the dataset variables show poor correlations displayed by reddish colors in the heat map. Such correlations between the data features indicate that regression analysis would not be able to sufficiently distinguish the complex relationship among features and labels, thereby justifying ML/DL approaches to achieve improved fluid-loss classification.

7.5 One hot encoding

There are multiple labels for prediction in the lost circulation classification tasks. To facilitate label detection by the algorithm, a matrix with a binary vector was assigned to each of the five lost circulation classes distinguished. This was performed during the training iterations. The algorithm distinguishes each label when it encounters the digit “1” in the vector. This approach is called One Hot Encoding and is widely used for word embedding in RNN architectures. In this work, we defined a vector consisting of five arrays, one for each lost circulation class label.

7.6 Performance measurement

In data mining, algorithms are structures containing data and their performance in the output prediction is evaluated based on different metrics. In this paper, the

performance of RNNs and CNNs were measured using the following performance assessment metrics: Accuracy, Precision, Recall, F-1 Score, and Confusion Matrix. Each cases of classification results have four possible outputs. True positive (TP) and true negative (TN) are two possible outputs that represent the positive and negative labels that have been correctly predicted. The other two possible outputs, i.e., false negative and false positive, denoted by FN and FP, respectively, are negative and positive outputs that have been incorrectly predicted. Equations (2), (3), (4) and (5) define the Accuracy, Precision, Recall and F-1 Score calculations, respectively, for classification, each involving just two outputs.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$F1\text{Score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{\text{Recall} + \text{Precision}} \quad (5)$$

For classification models with a multiclass dataset, it is essential to separately evaluate the performance of each model for all classes. The Confusion Matrix provides a simple display tool for assessing the performance of multi-feature classification. It considers an $n \times n$ matrix for n labels. In this matrix, the (i, j) -th array denotes the number of data inputs that are predicted incorrectly to belong to class i , while their true labels are located in class j [59].

7.7 Model execution

Each data record of the compiled lost circulation dataset once normalized, is classified into five lost circulation classes. Classification of the lost circulation events involves extracting the required data from the database. According to the downhole loss severity thresholds, any lost circulation event falls into one of the following classes: 1) No Loss (0 bph); 2) Seepage Loss (less than 10 bph); 3) Partial Loss (between 10 and 100 bph); 4) Severe Loss (more than 100 bph); and 5) Complete Loss (no return of the drilling mud at the surface) [69]. The number of data records belonging to each class from the Azadegan dataset is listed in Table 3. It is apparent that 75.8% of the data records in this dataset are associated with no fluid losses.

The normalized and classified dataset is then split into subsets of data records to build, train and test the deep

Fig. 9 HeatMap correlation plot

learning models. The data records were spread across all ranges in the training and testing subsets using *Sci-kit learn* machine learning library in *Python* programming language. Data splitting of the 65,376 data records (rows in the network comprising depth-based features and labels) involved 75% being allocated to the training subset. The remaining 25% of data records were allocated to the testing subset and held independently of the model training processes. RNNs are used primarily to predict time-series data over specified time intervals. To adapt RNNs to loss severity classification task, the dataset was converted into a sequential format comprising a single list, row by row. This was done by setting the time-step control variable equal to number 1. To inhibit overfitting by the models, an early-stopping technique was employed in DL architectures. The early-stopping callback allows the network to run a specified number of epochs but to stop training once the training process triggers less reliable performance.

7.8 Application of recurrent neural networks (RNN)

7.8.1 Long short-term memory (LSTM)

In order to classify the types of loss circulation events encountered, an LSTM layer with 32 neurons was created followed by 0.5 dropouts to regularize the network by removing all nodes that do not contribute to learning [59]. The architecture was completed with two fully connected layers that had 100 and 6 neurons that applied RELU and

Table 3 Number of data records in different classes of fluid loss

Losses type	Rate of losses (BBL/HR)	Number of data records
No loss	0	49,545
Seepage	< 10	12,880
Partial	10–100	2647
Severe	> 100	270
Complete	No Return	34

SOFTMAX as their activation functions, respectively. The optimum hyperparameter selection was performed based on an iterative process. The building elements and the number of trainable parameters within the LSTM model are listed Table 4. Since the model has sequential form, output shape of each layer is a one-dimension vector and *None* refers to the second dimension which is not utilized by this model. Finally, the model was trained and tested with a 0.01 learning rate.

Trainable parameters referred to in Tables 4, 5 and 6 are the weight matrixes which are established in different ways depending on the layer structures employed by each deep learning network. These numbers are determined by the optimized architectures of the model components, particularly the shape of each layer and the optimization strategy adopted by each model.

7.8.2 Gated recurrent units (GRU)

In comparison with LSTM, GRU has a simpler structure and requires a shorter calculation time [61]. To evaluate the performance of GRU in predicting lost circulation classification, a network similar to the LSTM was built incorporating a 0.01 learning rate with 32 GRU nodes, a 0.5 dropout layer, and completed with two dense layers. The activation functions of dense layers were similar to those used in the LSTM model. The architecture of the GRU network is listed in Table 5. The GRU network requires less parameters for training the model, compared to LSTM. The GRU model was also executed with a 0.01 learning rate.

7.9 Application of convolutional neural network (CNN)

In order to construct the convolutional neural network, a one-dimensional convolution layer consisting of 64 filters with a kernel size of one and RELU as the activation function was created. A max-pooling layer with a pooling size of one was used for subsampling the input date. In the next step, a flat layer was created to transform the convolutional output into a format that was suitable for fully connected layers. To finalize the architecture, two dense

layers using RELU and SOFTMAX activation functions with 50 and 6 neurons, respectively, were implemented. A drop-out layer was created between the two dense layers to regularize the flow of information through the model. The architecture of CNN network is listed in Table 6. The model was executed with a 0.001 learning rate.

8 Results and discussion

The performance of the DL approach in the classification of downhole lost circulation is evaluated through building and evaluating LSTM, GRU, and CNN models. These three models were executed using a large sequential dataset of drilling parameters, and mud rheology data (65,376 data records each with 17 influencing variables including: hole section, depth, rate of penetration, weight on bit, rotation, torque, standpipe pressure, flow in, flow out, pump stroke, mud weight, funnel viscosity, plastic viscosity, yield point, gel strength 10 s, gel strength 10 min, solid percentage). In order to compare the models in terms of speed and accuracy 100 iterations were performed as part of the execution sequence for each model run.

To compare the classification results of the three DL models, four prediction metrics were calculated: precision, recall, accuracy, and F1-score. Figure 10a–c and Fig. 11a–c reveals that none of the models stopped during model execution as the drop-out configuration prevented the early stopping mechanism from being triggered. Also, all

Table 4 LSTM network design

Layer type	Output shape	No. of trainable parameters
LSTM	(None, 32)	6400
Dropout	(None, 32)	0
Dense	(None, 100)	3300
Dense	(None, 6)	606

Table 5 GRU network design

Layer type	Output shape	No. of trainable parameters
GRU	(None, 32)	4800
Dropout	(None, 32)	0
Dense	(None, 100)	3300
Dense	(None, 6)	606

Table 6 CNN network design

Layer type	Output shape	No. of trainable parameters
Conv1d	(None, 1, 64)	1152
Max-pooling1d	(None, 1, 64)	0
Flatten	(None, 64)	0
Dense	(None, 50)	3250
Dropout	(None, 50)	0
Dense	(None, 6)	306

models have obtained high prediction accuracy associated with both training and testing data subsets. In addition, higher accuracy is achieved with the testing subsets than the training subsets indicating that the trained models are not substantially overfitting the dataset evaluated. The cross-entropy loss function was computed for each of the three models, as shown in Figs. 10a–c. The accuracies (Eq. (2)) achieved by the two RNN models (LSTM and GRU) was 0.94 (see Fig. 11a, b), confirming the capability of the memory states for adequately solving the lost circulation classification problem with a dataset of this size. The CNN model achieved even higher accuracy of 0.98. The convolution training and testing trends are illustrated in Figs. 10c, 11c revealing that the learning pattern of CNN model is smoother than that displayed by the two RNN models. As should be expected, the DL testing trends show more fluctuations than their training curves. Figure 10a–c illustrates that for the training and testing subsets the cross-entropy loss values decreased progressively as the number of epochs increased. This confirms that the network architectures kept learning from the data continuously throughout the training process. The training models appear not to have overfitted the training subset because the difference between two loss curves (testing and training) are very small for each model.

The three models have been executed in TensorFlow version 2.0.0 and run in a system with the following configuration: Windows 10, 2.67 GHz Corei5 processor, 4 GB installed memory (RAM), and 64-bit operating system. Although the GRU model required less time (average training speed: 5 s/step) to execute compared to the LSTM model (average training speed: 6 s/step), due to its underlying architecture, the CNN model consumed the least time (Average training speed: 4 s/step) for classification in each epoch (iteration). This outcome highlights the impressive speed of convolutional networks in performing classification tasks.

The models were separately evaluated for lost circulation classification using the defined range of accuracy metrics (Eq. (2), (3), (4) and (5)) with the results listed in Tables 7, 8 and 9. The dataset contained relatively few data records associated with the “complete loss” circulation state (only 34 out of 65,376, Table 3), compared to other classes. The RNN models were more successful in correctly classifying of the “complete loss” state than that the CNN model. On the contrary, when it came to classifying those lost circulation classes involving large numbers of data records, the CNN model provided superior prediction performance to the RNN models. For instance, the dataset included many more examples with the “seepage loss” class (12,880 of 65,376 data records, Table 3), for which the precision level (Eq. (3)) achieved by CNN was 0.94, while RNNs achieved the much lower precision of 0.80.

Fig. 10 **a** Categorical cross-entropy loss function curves for LSTM architectures calculated using Eq. (17). **b** Categorical cross-entropy loss function curves for GRU architectures calculated using Eq. (17). **c** Categorical cross-entropy loss function curves for CNN architectures calculated using Eq. (17)

Fig. 11 **a** Accuracy curves for LSTM architectures calculated using Eq. (2). **b** Accuracy curves for GRU architectures calculated using Eq. (2). **c** Accuracy curves for CNN architectures calculated using Eq. (2)

A confusion matrix was constructed based on the classification results for each model. The elements on the main diagonal of the confusion matrix represent the number of data points correctly classified for each defined lost circulation class. In Fig. 12a–c, the parts of dataset representing the numbers of data records correctly classified in the “no loss” and “seepage loss” classes are depicted in bold colors. The CNN method was able to correctly classify more data points with “no loss” and “seepage loss” labels (Fig. 12c). In contrast, the LSTM and GRU methods were able to correctly classify fewer data records from those two lost circulation classes (see Fig. 12a, b).

Figure 13a, b indicate the overall accuracy and loss function curves applied to training and testing subsets of data records for different models. The high accuracy and low loss function value obtained by the three models highlights the ability of the three DL models to accurately perform lost circulation classification in the large compiled subsurface data set for the Azadegan oil field.

9 Conclusions

The complexity of downhole lost circulation diagnosis and classification is a challenging and high-profile issue for the oil and gas drilling industry. Reducing its impact is vital for the efficient recovery of oil resources from subsurface

Table 7 Performance measurements achieved by the LSTM model for the testing subset of data records

Label	Metric type		
	Precision	Recall	F1-Score
Complete	1	0.8	0.89
No loss	0.98	0.94	0.96
Partial	0.74	0.89	0.81
Seepage	0.80	0.88	0.84
Severe	0.78	0.86	0.82

Table 8 Performance measurements achieved by the GRU model for the testing subset of data records

LABEL	Metric type		
	Precision	Recall	F1-score
Complete	1	0.40	0.57
No loss	0.98	0.94	0.96
Partial	0.75	0.88	0.81
Seepage	0.80	0.87	0.84
Severe	0.76	0.85	0.80

Table 9 Performance measurements achieved by the CNN model for the testing subset of data records

Label	Metric type		
	Precision	Recall	F1-score
Complete	0.88	0.70	0.78
No loss	0.99	0.98	0.99
Partial	0.75	0.95	0.84
Seepage	0.94	0.90	0.92
Severe	0.85	0.94	0.89

reservoirs. In the past decade, various studies have focused on implementing advanced methods addressing downhole problems. Several studies have applied conventional machine learning (ML) algorithms to predict lost circulation events. Although some ML methods performs

classification tasks with a certain degree of precision, classifying large datasets requires the application of more robust and efficient methods. Traditional ML tend to be constrained by the size of the datasets they can meaningfully process. Available ML algorithms are not able to achieve reliable predictions for field cases with very large quantities of data. In this study, a very large well dataset from Azadegan oilfield, Iran, is evaluated. The Sarvak, Gadvan, Fahliyan, and Kazhdumi formations constitute the field's carbonate oil reservoirs, and the wells drilled into those zones provide 17 parameters used to classify five downhole loss-event categories using three deep learning (DL) models: Convolutional Neural Network (CNN); Gated Recurrent Unit (GRU); and, Long-Short Term Memory (LSTM). For this field, the primary objective is to classify the stages of the lost circulation phenomenon in vuggy formations with carbonate and dolomitic lithologies, so

Fig. 12 Confusion matrices for: **a** LSTM, **b** GRU, **c** CNN models **Fig. 13** Comparison summary of deep learning models applied to the Azadegan oilfield data set: **a** Loss function curves, **b** Accuracy curves

as to accurately classify its severity. By doing so it places the drilling team in a better position to plan and initiate treatment actions before encountering drilling fluid losses. The ability to do this reduces drilling costs and time and improves resource recovery efficiency. Analysis of the deep learning models developed establish the following:

- Pre-processing of dataset, which is a key of DL architecture success, assists prediction performance by filtering out the noisy data points and outliers. The cleansed data is subsequently normalized prior to model processing to minimize bias towards certain input variables.
- The LSTM, GRU, and CNN models all achieved high prediction accuracies for lost circulation severity. However, the CNN architecture outperformed the other two methods.
- Early-stopping and drop-out configurations helped the DL models to avoid overfitting and thereby improved their generalization capabilities.
- The presence of more data records in the database tends to improve prediction accuracy in classification problem.
- Confusion matrices reveal that the accuracies achieved for classifying the “no loss” and “seepage loss” data records, which are more abundant in the the dataset studied, are higher than those achieved for other fluid-loss classes.
- The DL models developed achieve high accuracy by selective feature selection without using dimension reduction. This is possible because these models are designed to automatically ignore irrelevant features.
- The impressive performance of these DL networks with the Azadegan oilfield dataset suggest that the developed approach could be readily recalibrated to provide consistent predictions for potential lost circulation in other oil and gas fields. Further evaluation is required to confirm the models’ generalization capabilities.

Authors’ contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Sajjad Mardanirad], [David Wood] and [Hasan Zakeri]. The first draft of the manuscript was written by [Sajjad Mardanirad] and all authors contributed to the interpretation and commented on all versions of the manuscript. All authors read and approved the final manuscript.

Funding No funding was received to assist with the preparation of this manuscript.

Data and material availability The dataset evaluated in this study is available for readers to download as an Excel file.

Code availability The relevant code generated during the current study are available from the corresponding author on publisher request.

Declarations

Conflicts of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethics approval The authors have adopted a code of ethics in preparing and submitting this manuscript.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: Details of deep learning methods

The details provided in this appendix complement the summaries of the DL methods provided in Sect. 5.

Long short-term memory (LSTM)

After each iteration in the hidden layer, an internal memory is created that contains a copy of the last iteration's output that acts effectively as the memory of the cell. The information flow through an LSTM cell is controlled using three gates with different missions. The input gate, i , controls when and how much new and distant-past-iteration data should be remembered in the memory cell. On the other hand, f stands for the forget gate which controls when and how much past information is forgotten, g stands for the candidate internal memory of this time step which multiplies the input gate (i_t) by the hyperbolic tangent ($\tanh$) activation function [70] as part of Eq. (11), and o , represents the output gate where the remembered information is allowed to flow out from the memory cell to contribute to the current (t) iteration [71].

Eqs. (6), (7), (8), (9) and (10) describe the flow of information through an LSTM cell [61].

Firstly, the update rule of input gate is established, typically in the form of Eq. (6):

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (6)$$

Here, h_{t-1} stands for the last time step output, x_t denotes the current time step input data, c_{t-1} is the value of the last time step of cell memory, b_i represents the bias term of the input gate, W_{xi} , W_{hi} and W_{ci} indicate the weights related to the input, output and internal memory information of the present time step, and σ denotes the *sigmoid* activation function [70] respectively.

In a similar way, the forget gate adjustment rule is described using Eq. (7):

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (7)$$

The internal memory of this time step is calculated with Eq. (8):

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (8)$$

where f_t and i_t have been previously calculated. W_{xc} and W_{hc} represent the weight values of the internal memory of this time step for input and output information, and b_c denotes the bias term of internal memory parameter.

Once the memory state variable has been updated for the present iteration, the output gate value is obtained with Eq. (9):

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (9)$$

Finally, using the output gate values derived from the internal memory, the output variable value for this LSTM cell for the present iteration (t) is calculated with Eq. (10):

$$h_t = o_t \tanh(c_t) \quad (10)$$

The sigmoid activation function is implemented in LSTM gate value equations and returns a binary output that controls whether the information flows through the memory cell or not, i.e., is it forgotten or stored / processed. Furthermore, the hyperbolic tangent function ($\tanh$) influences the calculation of the internal memory value by ensuring it falls between zero and one.

Gated recurrent unit (GRU)

Equation (11) describes the update gate value of a GRU cell in prevailing time step (t):

$$z_t = \sigma_1(W_{hz}h_{t-1} + W_{xz}x_t) \quad (11)$$

where, h_{t-1} represents the hidden state of the prior time step, W_{hz} denotes the weight applied to input information from iteration (t-1), and W_{xz} represents the weight applied

to input information for the present iteration (t) or prevailing time step.

Similarly, the reset gate is written as Eq. (12):

$$r_t = \sigma_1(W_{hr}h_{t-1} + W_{xr}x_t) \quad (12)$$

where W_{hr} and W_{xr} are the weights applied to the input information from iteration (t-1) and for iteration (t) for the reset gate in this time step.

The linear interpolation of the previous and candidate hidden states yield the hidden state of the prevailing time step (t) by applying Eq. (13):

$$h_t = (1-z_t)h_{t-1} + z_t \tilde{h}_t \quad (13)$$

where the candidate hidden state is calculated using Eq. (14):

$$\tilde{h}_t = \sigma_2(W_{chx}x_t + W_{chr}(r_t \cdot h_{t-1})) \quad (14)$$

In Eq. (14), W_{chx} stands for the weight of the candidate hidden state input, and W_{chr} is the weight for the reset gate of the candidate hidden state in the prevailing time step (t).

Convolutional neural network (CNN)

The input layer of the CNN, which could potentially be an image data or a sequential time-series data set, processes that data by firstly applying filters and breaking it up into small, labelled fragments of data records as it is imported [62]. In the convolution layer, data are further filtered and a feature map is constructed by applying a *RELU* activation function [70] that is responsible for nonlinearity transformation of the data on to this feature map. The filter, referred to as a kernel, is a matrix of numerical values that are multiplied by the input data, in a similar way that weights are applied in conventional neural network architectures. The number of times a filter is specified by a CNN hyperparameter. The “pooling” layer then reduces the dimensionality of the feature map and divides it into further subsamples for more exhaustive analysis. The pooled and convoluted data then needs to be “flattened” to prepare it for entry into the logit layers. The logit layers, consisting typically of more than one fully connected dense layers, performs classification on the processed data and predicts the dependent variable (in this study lost circulation) class for each data record.

The one-dimension convolution operation at time t, $S_{(t)}$, is described by Eq. (15), which generates the feature map as an output [72]:

$$S_{(t)} = (X \times W)_{(t)} \quad (15)$$

where X refers to the input data and W describes the kernel matrix. If a multidimensional dataset in the axis of i and j are available, the convolution operation ($S_{(ij)}$) can be written in the form of Eq. (16):

$$S_{(ij)} = (K \times I)_{(ij)} = \sum_m \sum_n I_{(i-m, j-n)} K_{(m, n)} \quad (16)$$

where, K and I represent the kernel and input for a two-dimensional dataset, while m and n indicate the counts of input and output data records, respectively.

The categorical cross-entropy loss function [70], $L_{(c,p)}$, is mathematically described as Eq. (17):

$$L_{(c,p)} = - \sum_i C_i \ln(P)_i \quad (17)$$

where c stands for the true class while the predicted value is y , and p denotes the probability distribution of i -th observed value.

References

- Alkinani H, Alhameedi T, Dun-Norman S (2020) Data-driven decision-making for lost circulation treatments: a machine learning approach. *J Energy AI*. <https://doi.org/10.1016/j.egyai.2020.100031>
- Alkinani H, Alhameedi A, Dunn-Norman S, Alkhamis M, Mutar R (2019) Prediction of lost circulation prior to drilling for induced fractures formations using artificial neural networks. *SPE Oklahoma City Oil and Gas Symposium, Oklahoma City, Oklahoma, USA*. <https://doi.org/10.2118/195197-MS>
- Nayberg TM, Petty BR (1987) Laboratory study of lost circulation materials for use in oil-base drilling muds. *SPE Drill Eng* 2(03):229–236. <https://doi.org/10.2118/14723-PA>
- Howard GC, Scott PP Jr (1951) An analysis and the control of lost circulation. *J Pet Technol* 3(06):171–182. <https://doi.org/10.2118/951171-G>
- Salehi S, Nygaard R (2011) Evaluation of new drilling approach for widening operational window: implications for wellbore strengthening. *SPE Prod Op Symp, Okla, USA*. <https://doi.org/10.2118/140753-MS>
- Transparency Market Research. Drilling fluids market (2017) (Type: water-based muds, oilbased muds, and synthetic-based muds; and application: onshore and offshore)—global industry analysis, size, share, growth, trends, and forecast 2018–2026. From <https://www.transparencymarketresearch.com/drilling-fluid-market.html>.
- Caenn R, Darley HCH, Gray GR (2011) Composition and properties of drilling and completion fluids. *Gulf Prof Publ*. <https://doi.org/10.1016/C2009-0-64504-9>
- TensorFlow (2020) Install TensorFlow with pip. From <https://www.tensorflow.org/install/pip?hl=en/>.
- Parn-anurak S, Engler TW (2005) Modeling of fluid filtration and near-wellbore damage along a horizontal well. *J Petrol Sci Eng* 46:149–160. <https://doi.org/10.1016/j.petrol.2004.12.003>
- Sanfillipo F, Brignoli M, Santarelli FJ, Bezzola C (1997) Characterization of Conductive Fractures While Drilling. *SPE Eur Form Damage Conf Hague Neth*. <https://doi.org/10.2118/38177-MS>
- Lietard O, Unwin T, Guillot D, Hodder M (1996) Fracture width LWD and drilling mud / LCM selection guidelines in naturally fractured reservoirs. *Eur Pet Conf Milan Italy*. <https://doi.org/10.2118/36832-MS>
- Lavrov A, Tronvoll J (2004) Modeling mud loss in fractured formations. *Abu Dhabi Int Conf Exhib Abu Dhabi UAE*. <https://doi.org/10.2118/88700-MS>
- Lavrov A, Tronvoll J (2005) Mechanics of borehole ballooning in naturally-fractured formations. *SPE Middle East Oil Gas Show Conf Kingd Bahrain*. <https://doi.org/10.2118/93747-MS>
- Lavrov A, Tronvoll J (2006) Numerical analysis of radial flow in a natural fracture: applications in drilling performance and reservoir characterization. *Abu Dhabi International Petroleum Exhibition and Conference, 5–8 November, Abu Dhabi, UAE*. Doi: <https://doi.org/10.2118/103564-MS>.
- Shahri MP, Zeyghami M, Majidi R (2012) Investigation of fracture ballooning-breathing using an exponential deformation law and Herschel–Bulkley fluid model. *Spec Top Rev Porous Media Int J* 3(4):341–351
- Shahri MP, Mehrabi M (2012) A new approach in modeling of fracture ballooning in naturally fractured reservoirs. *SPE Kuwait Int Pet Conf Exhib Kuwait City Kuwait*. <https://doi.org/10.2118/163382-MS>
- Ozdemirtas M, Babadagli T, Kuru E (2009) Experimental and numerical investigations of borehole ballooning in rough fractures. *SPE Drill Complet* 24(2):P256–265. <https://doi.org/10.2118/110121-PA>
- Lavrov A (2013) Numerical modeling of steady-state flow of a non-newtonian power-law fluid in a rough-walled fracture. *Comput Geotech* 50:P101–109
- Majidi R, Miska S, Thompson L, Yu M, Zhang J (2010) Quantitative analysis of mud losses in naturally fractured reservoirs: the effect of rheology. *SPE Drill Complet* 25(4):P509–517. <https://doi.org/10.2118/114130-PA>
- Mohaghegh, SH.D. (2015). System and Method Providing Real-Time assistance to drilling operation. Patent.google.com/patent/US20150300151.
- Hedge C, Grey KE (2017) Use of machine learning and data analytics to increase drilling efficiency for nearby wells. *J Nat Gas Sci Eng* 40:P327–335. <https://doi.org/10.1016/j.jngse.2017.02.019>
- Hedge C, Daigle H, Grey KE (2018) Performance comparison of algorithms for real-time rate of penetration optimization in drilling using data-driven models. *SPE J*. <https://doi.org/10.2118/191141-PA>
- Dunlop J, Isangulov R, Aldred WD, Sanchez HA, Flores JLS, Heidoiza JA, Belaskie J, Luppens C (2011) Increased Rate of Penetration Through Automation. *SPE/IADC Drill Conf Exhib Amst Neth*. <https://doi.org/10.2118/139897-MS>
- Yavari H, Sabah M, Khosravanian R, Wood DA (2018) Application of adaptive neuro-fuzzy inference system and mathematical ROP models for prediction of drilling rate. *Iran J Oil Gas Sci Tech* 7(3):73–100
- Sabah M, Talebkeikhah M, Wood DA, Khosravanian R, Ane-mangely M, Younesi A (2019) A machine learning approach to predict drilling rate using petrophysical and mud logging data. *Earth Sci Info*. <https://doi.org/10.1007/s12145-019-00381-4>
- Castiniera D, Toronyi R, Saleri N (2018) Machine learning and natural language processing for automated analysis of drilling and completion data. *SPE Kingd Saudi Arabia Annu Tech Symp Exhib Held Dammam, Saudi Arabia*. <https://doi.org/10.2118/192280-MS>
- Kowalchuk P (2019) Implementing a drilling reporting data mining tool using natural language processing sentiment analysis techniques. *SPE Middle East Oil and Gas Show and Conference, 18–21 March, Manama, Bahrain*. Doi:<https://doi.org/10.2118/194961-MS>.
- Fatehi M, Asadi H (2017) Data integration modeling applied to drill hole planning through semi-supervised learning: a case

- study from the Dalli Cu-Au porphyry deposit in the central Iran. *J Afr Earth Sci* 128:P147–160. <https://doi.org/10.1016/j.jafrearsci.2016.09.007>
29. Pollock, J., Stoecker-Sylvia, Z., Veedu, V., Panchal, N., Elshahawi, H. (2018). Machine Learning for Improved Directional Drilling. *Offshore Technology Conference, 30 April - 3 May, Houston, Texas, USA*. <https://doi.org/10.4043/28633-MS>.
 30. Zhang L, Wu S, Zheng W, Fan J (2018) A dynamic and quantitative risk assessment method with uncertainties for offshore managed pressure drilling phases. *Saf Sci* 104:39–54. <https://doi.org/10.1016/j.ssci.2017.12.033>
 31. Sule I, Khan F, Butt S, Yang M (2018) Kick control reliability analysis of managed pressure drilling operation. *J Loss Prev Process Ind* 52:7–20. <https://doi.org/10.1016/j.jlp.2018.01.007>
 32. Bhandari J, Abbassi R, Garaniya V, Khan F (2015) Risk analysis of deepwater drilling operations using Bayesian network. *J Loss Prev Process Ind* 38:P11–23. <https://doi.org/10.1016/j.jlp.2015.08.004>
 33. Hoffman J, Mao Y, Wesley A, Taylor A (2018) Sequence mining and pattern analysis in drilling reports with deep natural language processing. *SPE Annual Technical Conference and Exhibition, 24–26 September, Dallas, Texas, USA*. Doi: <https://doi.org/10.2118/191505-MS>.
 34. Alhameedi AT, Dunn-Norman S, Alkinani HH, Flori RE, Hilgedick SA (2017) Limiting drilling parameters to control mud losses in the dammam formation, South Rumaila Field, Iraq. *51st US Rock Mechanics / Geomechanics Symposium, USA, ARMA-2017–0930*
 35. Alhameedi AT, Alkinani HH, Dunn-Norman S, Flori RE, Hilgedick SA, Amer SA, Alsaba M (2018) Mud loss estimation using machine learning approach. *J Petrol Explor Prod Tech* 9:P1339–1354. <https://doi.org/10.1007/s13202-018-0581-x>
 36. Toreifi H, Rostami H, Khaksar A (2014) New method for prediction and solving the problem of drilling fluid loss using modular neural network and particle swarm optimization algorithm. *J Petrol Explor Prod Tech* 4:371–379. <https://doi.org/10.1007/s13202-014-0102-5>
 37. Cristofaro RA, Longhin GC, Waldmann AA, De Sá CHM, Vadinal RB, Gonzaga KA, Martins AL (2017) Artificial intelligence strategy minimizes lost circulation non-productive time in Brazilian deep-water pre-salt. *Offshore Technol Conf Brasil held Rio de Janeiro Brazil*. <https://doi.org/10.4043/28034-MS>
 38. Behnoudfar P, Hosseini P (2016) Estimation of lost circulation amount occurs during under balanced drilling using drilling data and neural network. *Egypt Petrol Res Inst* 26:P627–634. <https://doi.org/10.1016/j.ejpe.2016.09.004>
 39. Khaksar A, Rostami H, Niknafs H (2017) Integrated lost circulation prediction in oil field drilling operation, heavy oil. Nova Science Publishers Inc, NY, USA, pp 243–253
 40. Sabah M, Talebkeikhah M, Agin F, Talebkeikhah F, Hasheminasab E (2019) Application of decision tree, artificial neural networks, and adaptive neuro-fuzzy inference system on predicting lost circulation: a case study from marun oil field. *J Petrol Sci Eng* 177:P236–249. <https://doi.org/10.1016/j.petrol.2019.02.045>
 41. Geng Z, Wang H, Fan M, Lu Y, Nie Z, Ding Y, Chen M (2019) Predicting seismic based risk of lost circulation using machine learning. *J Petrol Sci Eng*. <https://doi.org/10.1016/j.petrol.2019.01.089>
 42. Alkinani HH, Alhameedi ATT, Dunn-Norman, Alalwani M, Mutar RA, AlBazzaz WH (2019) Data-driven neural network model to predict equivalent circulation density ECD. *SPE Gas & Oil Technology Showcase and Conference, Dubai, UAE*. <https://doi.org/10.2118/198612-MS>
 43. Li Z, Chen M, Jin Y, Lu Y, Wang H, Geng Z, Wei S (2018) Study on intelligent prediction for risk level of lost circulation while drilling based on machine learning. In: *52nd U.S. Rock Mechanics/ Geomechanics Symposium, 17–20 June, Seattle, Washington, USA, ARMA-2018–105*.
 44. Ahmed A, Elkhatatny S, Abdulraheem A, Abughaban M (2019) Prediction of lost circulation zones using support vector machine and radial basis function. *Int Pet Technol Conf, Dhahran, Kingd Saudi Arabia*. <https://doi.org/10.2523/IPTC-19628-MS>
 45. Hou X, Yang J, Yin Q, Liu H, Zheng J, Wang J, Cao B, Zhao X, Hao M, Liu X (2020) Lost Circulation prediction in South China Sea using machine learning and big data technology. *Offshore Technol Conf, Houston, USA*. <https://doi.org/10.4043/30653-MS>
 46. Jahanbakhshi R, Keshavarzi R (2014) Quantitative and qualitative analysis of lost circulation in natural and induced fractured formations: the integration of operational conditions and geomechanical parameters. *Eur J Environ Civ Eng*. <https://doi.org/10.1080/19648189.2014.949872>
 47. Wang Ch, Liu G, Yang Zh, Li J, Zhang T, Jiang H, Cao Ch (2020) Downhole working conditions analysis and drilling complications detection method based on deep learning. *J Nat Gas Sci Eng*. <https://doi.org/10.1016/j.jngse.2020.103485>
 48. Aljubran M, Ramasamy J, Lbassam M, Magana-mora A (2021) Deep learning and time-series analysis for the early detection of lost circulation incidents during drilling operations. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2021.3082557>
 49. Sabah M, Mehrad M, Ashrafi SB, Wood DA, Fathi Sh (2021) Hybrid machine learning algorithms to enhance lost-circulation prediction and management in the Marun oil field. *J Petrol Sci Eng*. <https://doi.org/10.1016/j.petrol.2020.108125>
 50. Moazzeni A, Nabaei M, Azari A (2011) Reducing consumed energy while drilling an oil well through a deep rig time analysis. *Adv Pet Explor Dev* 1(1):22–31
 51. Liang H, Yongqiang T, Xiang L, Yangyang L (2014) Research on drilling kick and loss monitoring method based on Bayesian Classification. *Pak J Stat Oper Res* 30(6):1251–1266
 52. Wu S, Zhang L, Zheng W, Liu Y, Lundteigen MA (2016) A DBN-based risk assessment model for prediction and diagnosis of offshore drilling incidents. *J Nat Gas Sci Eng* 34:139–158
 53. Agin F, Khosravanian R, Karimifard M, Jahanshahi A (2018) Application of adaptive neuro-fuzzy inference system and data mining approach to predict lost circulation using DOE technique (Case study: maroon oilfield) petroleum. *Southwest Petroleum University*. <https://doi.org/10.1016/j.petlm.2018.07.005>
 54. Wu S, Zhang L, Fan J, Zheng W, Zhou Y (2019) Real-time risk analysis method for diagnosis and warning of offshore downhole drilling incident. *J Loss Prev Process Ind* 62:103933
 55. Abbas AK, Al-haideri NA, Bashikh AA (2019) Implementing artificial neural networks and support vector machines to predict lost circulation. *Egypt J Pet* 28(4):339–347
 56. Osarogiagbon AU, Khan F, Venkatesan R, Gillard P (2020) Review and analysis of supervised machine learning algorithms for hazardous events in drilling operations. *Process Saf Environ Prot*. <https://doi.org/10.1016/j.psep.2020.09.038>
 57. GitHub, (2017, February 11). TensorFlow Release 1.0.0. from <https://github.com/tensorflow/tensorflow/blob/07bb8ea2379bd459832b23951fb20ec47f3fdbd4/RELEASE.md>.
 58. Wood DA (2020) Country-wide German hourly wind power dataset mined to provide insight to predictions and forecasts with optimized data-matching machine learning. *Renew Energy Focus* 34:69–90. <https://doi.org/10.1016/j.ref.2020.06.005>
 59. Ramsundar B, Bosagh Zadeh R (2018) TensorFlow for deep learning from linear regression to reinforcement learning, 1st edn. O'Reilly, USA
 60. Bahdanau D, Cho K, Bengio Y (2016) Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015, San Diego, USA*. <https://arxiv.org/abs/1409.0473v7>

61. Gao Y, Glowacka D (2016) Deep gate recurrent neural network. *JMLR Workshop Conf Proc* 63:350–365
62. Rungta K (2019) TensorFlow learn in 1 Day: make your own neural network. PublishDrive. https://books.google.de/books/about/TensorFlow_in_1_Day.html?id=Hb7jDwAAQBAJ&redir_esc=y
63. Shuka N (2017) Machine learning with TensorFlow, MEAP. Manning Publications, New York
64. Liu H, Guo R, Dong JC (2013) Productivity evaluation and influential factor analysis for Sarvak Reservoir in South Azadegan oil field Iran. *Petrol Explor Dev* 40(5):585–90
65. Du Y, Yi YJ, Xin J (2015) Genesis of large-amplitude tilting oil-water contact in Sarvak formation in South Azadegan Oilfield, Iran. *Pet Geol Exp* 37(2):187–193 **((in Chinese))**
66. Du Y, Xin J, Xu QC (2015) The rudist buildup depositional model based on reservoir architecture: a case from the Sarvak Reservoir of The SA oilfield, Iran. *Acta Sedimentol Sin* 33(6):1247–57 **((in Chinese))**
67. Zabihi R, Schaffie M, Nezamabadi-pour H, Ranjbar M (2011) Artificial neural network for permeability damage prediction due to sulfate scaling. *J Petrol Sci Eng* 78(3–4):575–581. <https://doi.org/10.1016/j.petrol.2011.08.007>
68. Saeedi A, Camarda KV, Liang J-T (2007) Using neural networks for candidate selection and well performance prediction in water-shutoff treatments using polymer gels—a field-case study. *Soc Pet Eng*. <https://doi.org/10.2118/101028-PA>
69. Nelson E, Guillot D (2006) Well cementing, 2nd edn. Schlumberger, UK, p 773
70. Gulli A, Kapoor A, Pal S (2019) Deep learning with TensorFlow 2 and keras, 2nd edn. Packt, Birmingham
71. GitHub (2019), Heaton, J. Deep Learning course note, Washington University, from https://github.com/jeffheaton/t81_558_deep_learning.
72. Goodfellow I, Bengio Y, Courville A (2015) Deep learning. MIT Press

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.