Effective prediction of lost circulation from multiple drilling variables: a class imbalance problem for machine and deep learning algorithms

David A. Wood¹ · Sajjad Mardanirad² · Hassan Zakeri³

Received: 23 September 2021 / Accepted: 30 November 2021 / Published online: 8 December 2021
© The Author(s) 2021

Abstract

Multiple machine learning (ML) and deep learning (DL) models are evaluated and their prediction performance compared in classifying five wellbore fluid-loss classes from a 20-well drilling dataset (Azadegan oil field, Iran). That dataset includes 65,376 data records with seventeen drilling variables. The dataset fluid-loss classes are heavily imbalanced (> 95% of data records belong to the less significant loss classes 1 and 2; only 0.05% of the data records belong to the complete-loss class 5). Class imbalance and the lack of high correlations between the drilling variables and fluid-loss classes pose challenges for ML/DL models. Tree-based and data matching ML algorithms outperform DL and regression-based ML algorithms in predicting the fluid-loss classes. Random forest (RF), after training and testing, makes only 35 prediction errors for all data records. Consideration of precision recall and F1-scores and expanded confusion matrices show that the RF model provides the best predictions for fluid-loss classes 1 to 3, but that for class 4 Adaboost (ADA) and class 5 decision tree (DT) outperform RF. This suggests that an ensemble of the fast to execute RF, ADA and DT models may be the best way to practically achieve reliable wellbore fluid-loss predictions. DL models underperform several ML models evaluated and are particularly poor at predicting the least represented classes 4 and 5. The DL models also require much longer execution times than the ML models, making them less attractive for field operations that require prompt information regarding rapid real-time decision responses to pending class-4 and class-5 fluid-loss events.

Keywords Drilling fluid loss of circulation · Imbalanced classification system · Expanded confusion matrices · Carbonate reservoirs · Fluid-loss severity classes · Precision versus recall

Introduction

Drill-fluid losses cause problems for many subsurface operations. They are often costly and time-consuming to mitigate. Early warning of such events, being aware of the appropriate treatment strategies suited to specific formations, and having supplies at hand of the necessary lost circulation materials

typically help drilling teams to react quickly to minimize their impacts. If not dealt with adequately, severe losses of drilling fluid from the well bore run the risk of the drill string becoming stuck and can potentially lead to blowouts, in addition to the high costs of replacing the lost drilling fluid volumes. Moreover, invasion of reservoir formations by drilling fluid and inappropriate loss-control materials can lead to formation damage and loss of long-term production (Van Der Zwaag 2006; Zhao et al. 2019; Alkinani et al. 2020a, b). The severity of drill-fluid losses in the subsurface varies and it can be usefully classified based on the measured rate of loss.

There are various factors that influence the occurrence of drilling fluid losses from an uncased wellbore. Conditions in which wellbore fluid pressure substantially exceeds a permeable formations fluid pressure are likely to result in fluid losses from the wellbore to the formation. Also, highly porous and permeable formations, particularly those impacted by fractures, faults and/or with large vuggy pore

✉ David A. Wood
dw@Dwasolutions.com
Sajjad Mardanirad
mardanirad1@gmail.com
Hassan Zakeri
hzakeri@pasargadep.com

¹ DWA Energy Limited, Lincoln, UK

² Oil Industries Engineering and Construction Group, Tehran, Iran

³ Pasargad Exploration and Production, Tehran, Iran

space (a characteristic of some carbonate zones), are particularly prone to wellbore fluid losses (Savari et al. 2016; Su et al. 2019). Figure 1 illustrates diagrammatically the type of subsurface formation conditions most at risk to substantial wellbore fluid losses.

Applying excessive wellbore fluid densities while drilling often induces fractures in the more brittle rock formations drilled that may exist in the subsurface with relatively sparse natural fractures. Such induced fractures are a common cause of wellbore fluid losses (Feng and Gray 2018; Alkinani et al. 2019a). In cases where wellbore fluid losses are minor it is often possible for drilling to proceed to a planned depth with the losses then being resolved by setting a casing string. However, with more substantial fluid losses immediate responses are required to safeguard well integrity. For some problematic zones it is possible to adopt underbalanced drilling techniques to reduce wellbore fluid losses (Abel and Gokhan 2004). However, in many situations the formation pressures of surrounding formation prevent the use of underbalanced drilling. Also, in severely depleted fractured reservoirs, wellbore fluid losses may still persist even when underbalanced drilling conditions are adopted.

Substantial wellbore fluid losses in non-reservoir formations can sometimes be cured by introducing a plug of cement or lost circulation material (LCM), as formation damage is typically not relevant in such zones. Substantial wellbore fluid losses into producing reservoir formation are, however, more difficult to resolve without risking damage to the reservoir formation. Various LCMs have been successfully applied in various drilling conditions to either cure or significantly reduce fluid losses while drilling. However, LCM tends to be more effective when they are tailored to specific subsurface conditions. However, in many drilling operations, when fluid losses occur, the type of LCM used is

either chosen by trial and error or governed by what LCM is available at the location (Salehi and Nygaard 2011).

In field development circumstances, where drilling information from multiple wells is available, statistical assessment, modelling and data mining of that information, particularly with the assistance of artificial intelligence methods, can provide useful insight. Such approaches can help to better determine the type of fluid losses occurring and what should be anticipated as drilling progresses. They can also help to identify the type and formulation of LCM that is most likely to resolve the problem (Alkinani et al. 2018, 2019b; Mardanirad et al. 2021).

Loss of wellbore fluids in the subsurface poses difficulties to model and predict from drilling variables, because there are a substantial number of variables involved that are often poorly correlated with each other or fluid-loss conditions, and those relationship tend to vary from one formation to another. Various attempts have been made to develop analytical models to predict wellbore fluid losses, including the application of Darcy's law to fluid filtration and formation permeability (Parn-Anurak and Engler 2005) and quantifying fracture characteristics in a range of porous media into fractured media using quantitative approaches (Sanfillippo et al. 1997; Shahri and Mehrabi 2012; Lavrov 2013). Some numerical solutions have applied the Herschel–Buckley general power law for non-Newtonian fluids to estimate wellbore fluid losses in fractured reservoirs (Majidi et al. 2010). Alhameedi et al. (2017) applied regression analysis to develop statistical formula relating a range of drilling variables considered to influence wellbore fluid losses to suggest controls that could be applied to drilling conditions in order to minimize lost circulation. The diversity of subsurface conditions, both within the wellbore and the formations penetrated, means that none of the proposed formulations can be relied upon, especially in vuggy fractured carbonate reservoirs.

Numerous, data-driven, artificial intelligence models, both machine learning (ML) and deep learning (DL) have been proposed over the past decade or so to predict wellbore fluid loss based on drilling variables. These have been recently described and summarized (Mardanirad et al. 2021). To be effective, it is apparent that data need to be gathered and assessed from multiple wells in a specific field area in order to establish confidence in ML/DL predictions generated. Cristofaro et al. (2017) compared the performance of several classifiers, including customized Naïve Bayes, K-nearest neighbour (KNN), radial basis function (RBF) and multi-layer perceptron (MLP) to determine optimum LCM solutions to apply offshore Brazil. They used that analysis to derive an instance-based interface combining several ensemble ML models, i.e. combining the results of several models into a single score. Hou et al. (2020) applied an MLP model to successfully predict wellbore fluid losses while drilling

Fig. 1 Subsurface conditions commonly associated with drilling fluid losses

a high-pressure high-temperature (HPHT) field offshore in the South China Sea.

Support vector machine/classifier (SVM/SVC) models are known to be potentially effective wellbore fluid-loss predictors. Behnoudfar and Hosseini (2016) applied an SVM model to predict fluid losses during underbalanced drilling in Egypt. Manshad et al. (2017) successfully applied an SVM model to a heavy oil field in Iran. Ahmed et al. (2019) applied an SVM model with an RBF kernel to predict lost circulation in two wells from Saudi Arabia with high accuracy. A least squares support vector machine (LSSVM) model optimized with a cuckoo search algorithm delivered better loss circulation prediction accuracy for the Marun oil field than MLP optimized models (Sabah et al. 2021). Jahanbakhshi and Keshavarzi (2015) included geomechanical factors, including in situ stresses and rock strengths, as well as wellbore drilling variables to predict lost circulation and demonstrated that SVM models outperformed MLP models with their dataset.

Adaptive neuro-fuzzy inference systems (ANFIS) have been compared with optimized MLPs and decision tree (DT) models (Marun field, Iran), showing that DT outperformed the other models for that dataset (Sabah et al. 2019). An ANFIS model was also applied to data from that field by Agin et al. (2020). Li et al. (2018) compared MLP, random forest (RF) and SVM models to predict lost circulation from drilling data from an Iraqi oil field, demonstrating that the RF model outperformed the other models. Magzoub et al. (2021) compared four models, RF, KNN, gradient boosting and Adaboost in assessing the relationship between drilling fluid components, rheology and lost circulation. They found that the gradient boosting model performed the best. Optimized data matching ML models, such as the transparent open box (TOB), which like KNN do not rely on variable correlation relationships, have also been applied to predict wellbore fluid loss from drilling parameters (Wood 2018).

In recent years, DL methods have been evaluated as wellbore fluid-loss prediction tools from larger dataset with some success. Wang et al. (2020) applied a one-dimensional convolutional neural network (CNN) to distinguish normal down-hole working conditions based on a time series from abnormal conditions. Mardanirad et al. (2021) compared the performance of CNN with gated recurrent units (GRU) and long short-term memory (LSTM) networks applied to a large 20-well dataset and established that their CNN model outperformed the other DL models in predicting classes of wellbore fluid losses. Aljubran et al. (2021) compared DL CNN and LSTM models for rapid detection of wellbore fluid losses with RF and MLP ML models using time series analysis and found the CNN model yielded the most accurate predictions.

An issue that has yet to be addressed in lost circulation prediction studies is the “*class imbalance*” problem (Guo

et al. 2008), i.e. where the distribution of data records in the classes to be predicted is strongly skewed, such that a large proportion of the records fall within certain classes and very few in other classes. It is a well-studied issue in classification ML modelling (Ling and Sheng 2011; Maheshwari et al. 2017) considered in many fields, and addressed with respect to some specific oil and gas related problems, such as gas pipelines (Choubineh et al. 2020) and enhanced oil recovery screening (Pirizadeh et al. 2021). When classifying lost circulation from multi-well drilling data into classes varying from “no loss” to “complete loss” the number of data records in the no/low-loss classes tend to be orders of magnitude more than in the severe/complete-loss classes. Hence, large loss circulation datasets tend to be highly imbalanced and that imbalance should be considered when assessing the prediction accuracy of ML/DL models. Most of previously published work that has predicted and classified lost circulation from drilling variables applying ML models has done so using relatively small sets of data (i.e. less than a few thousand data records). More recently, the DL models mentioned have been applied to large datasets. However, the work to date has largely ignored the class imbalance issue which becomes more significant in larger datasets.

The novelties of this study are that it considers a large, 20-well, dataset (65,376 data records), it evaluates and compares the wellbore fluid-loss class prediction accuracy using 11 distinct ML/DL methods (12 models evaluated) and takes into account the class imbalance issue with the aid of an expanded confusion matrix presentation not previously employed.

This work is organized as follows: “[Dataset description](#)” section characterizes the dataset for the Azadegan oil field and the fluid-loss classes defined; “[Prediction methods evaluated to predict loss circulation classes](#)” section describes the ML/DL methods evaluated and their control parameter configurations; “[Class prediction accuracy measures assessed](#)” section defines the classification prediction accuracy metrics assessed; “[Pre-processing and process workflow applied to the fluid-loss dataset](#)” section explains the pre-processing steps applied to the dataset and the general workflow adopted for the ML/DL model evaluations; “[Results](#)” section presents the study’s results; “[Discussion](#)” section discusses the results in terms of datasets class imbalance with the aid of expanded confusion matrices; and “[Conclusions](#)” section draws conclusions.

Dataset description

A large (65,376 data records from twenty wells), published, drilling parameter dataset (southern portion Azadegan field in southwest Iran) has been compiled (Mardanirad et al. 2021) and is used here to evaluate and predict drilling

fluid-loss conditions with multiple models. It is a giant heavy oil ($< 20^\circ\text{API}$ gravity) field that is only partially developed (Fig. 2 Map). It produces oil mainly from the Upper Cretaceous, Sarvak carbonate, reservoirs located at about 2700 m below surface (Du et al. 2016). However, there are other oil-bearing reservoirs and permeable zones present in the field (Fig. 3) linked to a prolific petroleum system (Liu et al. 2013). The field consists of a large domal structure with two crestal points (north and south) with the oil water contact sloping from south to north across the southern crestal portion of the field (Du et al. 2016). There are four oil producing zones within the Sarvak reservoir, of which the S-3 zone is the most prolific producer and reserves holder. Most drilling fluid-loss events are recorded in the S-3 zone.

The compiled data records are arranged in order of increasing depth, with variables drawn from available well reports from each well. Minor (< 10 bph) and no fluid losses are reported in certain zones of almost all formations penetrated by the wells. More substantial fluid losses (> 10 bph) tend to be concentrated in five specific formations beneath the Asmari, The lower Tertiary Pabdeh-Jahrum and Pabdeh formations, and the Cretaceous Gurpi, Sarvak and Gadvan formations. The most severe fluid losses (> 100 bph) have been recorded in the carbonate zones of the Jahrum, Pabdeh-Jahrum and Sarvak formations. The fractured Jahrum dolomites are particularly problematic in relation to fluid losses with many of the complete lost circulation events recorded in that formation.

The possible occurrence of fluid losses to varying degrees in all formations penetrated by drilling in the Azadegan field means that reliable systems that provide the rapid analysis of drilling variables to predict the type of fluid losses occurring are valuable. Armed with such analysis it is possible for

drilling teams to deploy appropriate mitigation strategies and select the most suitable lost circulation materials to apply for the type of losses observed. Artificial intelligence models based on drilling variables have the potential to provide such reliable systems. This study is designed to evaluate and compare the performance of machine and deep learning methods in predicting different classes of fluid losses from recorded drilling variables.

Table 1 provides a statistical summary of the drilling parameters recorded in the 20-well dataset evaluated to predict lost circulation conditions. The statistical details reveal that for the most part, the variables are symmetrically distributed and cover a wide range of conditions.

All dataset records are distinguished as belonging to one of five lost circulation categories (Table 2). The categories (classes) are defined based on generally accepted loss severity value ranges (Nelson and Guillot 2006). These are: no-loss class 1 (0 bph; barrels per hour); seepage-loss class 2 ($0 < \text{bph} < 10$); partial-loss class 3 ($10 \leq \text{bph} \leq 100$); severe-loss class 4 (> 100 bph); and complete-loss class 5 (no drilling fluid returns to the surface). The dataset is highly imbalanced with respect to the data record numbers associated with each fluid loss (Table 2). Indeed more than 95% of the data records belong to classes 1 and 2, whereas $< 0.5\%$ belong to classes 4 and 5. Although such a distribution is to be expected from typical drilling operations, it poses substantial challenges to the prediction algorithms in terms of accurately distinguishing the relatively sparse data records belonging to fluid-loss classes 4 and 5.

Figure 4 displays a correlation coefficient matrix among the dataset variables in the form of a heat map. This reveals that the fluid-loss classes (dependent variable) is not well correlated with any of the input variables.

Fig. 2 Azadegan oil field location maps

Fig. 3 Stratigraphic section encountered in the Azadegan oil field. Modified after Du et al. (2016)

The best correlations that exist are between loss class and MD (+0.28), GS10s (−0.16), GS10m (−0.15), and TOQ (+0.14). The general lack of moderate to high correlations between loss classes and input variables poses a challenge for those prediction algorithms that rely particularly on the variable correlation relationships.

Prediction methods evaluated to predict loss circulation classes

The five classes of drilling fluid conditions, varying from “no loss” to “complete loss”, distinguished in the

Table 1 Statistical distributions of the eighteen variables considered in the dataset compiled from twenty wells in the Azadegan oil field

Statistical summary of eighteen variables for 65,376 data records constituting the Azadegan oil field drilling lost circulation dataset																	
Variable name	Variable unit	Variable abbreviation	Minimum	Maximum	Range	Average	Standard deviation	Standard error	Coefficient of variation	Percentile (%)							
										5	10	25	50	75	90	95	
Measured depth	(m)	MD	14	4285	4271	2046.3	1048.4	4.1003	0.5123	340	558	1172	2098	2905	3437	3654	
Rate of penetration	(m/hr)	ROP	0.1	616.0	616.0	10.6	13.7	0.0537	1.2913	1.6	2.1	3.6	6.4	12.3	23.6	33.3	
Weight on bit	(klbs)	WOB	0	79.4	79.4	15.7	11.1	0.0434	0.7058	2.7	4.1	7.5	12.7	21.2	33	38.9	
Rotation	(RPM)	ROT	0	457	457	160.6	48.62	0.1902	0.3027	85	100	130	155	186	230	246	
Torque	(klbs-ft)	TOQ	0	12.3	12.3	3.3	1.72	0.0067	0.5243	0.0	0.4	2.4	3.3	4.3	5.3	6.0	
Standpipe pressure	(psi)	SP	9	3915	3906	2047.0	594.4	2.3246	0.2904	953	1192	1661	2137	2542	2749	2823	
Flow in rate	(gpm)	Fin	0	1934	1934	592.6	222.0	0.8684	0.3747	211	227	478	593	782	898	920	
Flow out rate	(%)	Fout	0	100.1	100.1	59.9	15.6	0.0608	0.2598	36	41	50	58	72	82	86	
Pump strokes	(spm)	STK	6	242	236	112	37	0.1447	0.3298	45	49	95	115	141	158	164	
Mud weight	(pcf)	MW	66	127	61	79.3	10.9	0.0428	0.1381	70	72	75	77	80	82	116	
Funnel viscosity	(sec/qt)	Fvis	29	76	47	43.0	5.66	0.0221	0.1316	35	36	39	43	45	50	54	
Plastic viscosity	(cp)	Pvis	3	48	45	14.0	6.73	0.0263	0.4792	6	7	10	13	16	20	29	
Yield point	(lb/100ft ²)	YP	2	47	45	16.1	4.73	0.0185	0.2941	10	11	13	15	19	22	25	
Gel strength 10 s	(lb/100ft ²)	GS10s	1	26	25	4.4	2.27	0.0089	0.5209	2	3	3	4	5	7	9	
Gel strength 10 min	(lb/100ft ²)	GS10m	1	27	26	5.9	2.74	0.0107	0.4661	3	4	4	5	7	8	10	
Solid	(%)	SLD	4	65	61	14.3	6.04	0.0236	0.4213	8	9	11	13	16	18	31	
Hole section diameter	(inch)	HD	4	26	22	11.6	4.51	0.0176	0.3886	6.1	6.1	8.5	8.5	17.5	17.5	17.5	
Losses																	
(Dependent variable)	Classes 1 to 5	Loss	1	5	4	1.3	0.56	0.0022	0.4369	1.0	1.0	1.0	1.0	1.0	2.0	2.0	

Table 2 Distribution of the 65,376 data records among the fluid-loss classes

Fluid-loss class	Fluid-loss rate (bph)	Data records/class
No losses (1)	0	49,545 (75.79%)
Seepage losses (2)	0 < bph < 10	12,880 (19.70%)
Partial losses (3)	10 ≤ bph ≤ 100	2,647 (4.05%)
Severe losses (4)	> 100	270 (0.41%)
Complete loss (5)	No fluid returns	34 (0.05%)

Azadegan field dataset are predicted using eight widely applied machine learning (ML) and three deep learning (DL) methods. These methods are summarized in Table 3.

These methods are based on distinct mathematical algorithms, mainly involving regression techniques capable of

handling complex and nonlinear relationships between the input variables and dependent variable. The KNN algorithm is distinctive in that it employs data matching rather than correlations and/or regression to formulate its predictions. The algorithm codes for these DL/ML are executed in Python and based on customized versions published packages. The ML algorithms (ADA, DT, KNN, MLP, NBC, QDA, RF and SVR) are executed with Scikit-Learn algorithm functions (SciKit Learn 2021). The deep learning algorithms (CNN, GRU and LSTM) are executed with TensorFlow/Keras functions (TensorFlow 2021). Two versions of the CNN algorithm, with distinctive structures, are evaluated resulting in twelve models being considered in total. This study focuses on customized applications of these established ML/DL algorithms. Therefore, the mathematical formulas involved in the execution of each algorithm are not

Pearson's Correlation Coefficient Matrix for Eighteen Variables that Constitute the Azadegan Oil Field Drilling Lost Circulation Dataset																		
(R)	MD	ROP	WOB	ROT	TOQ	SP	Fin	Fout	STK	MW	Fvis	Pvis	YP	GS10s	GS10m	SLD	HD	Loss
MD	1.0000	-0.3312	-0.4753	0.4767	-0.0526	0.5496	-0.7957	-0.1680	-0.7400	0.0484	0.3253	0.3510	0.2456	-0.3695	-0.3444	0.1409	-0.8612	0.2796
ROP		1.0000	0.0112	-0.0393	0.0727	-0.2616	0.2087	0.0263	0.2017	-0.1788	-0.2803	-0.2632	-0.1938	0.0236	0.0199	-0.2197	0.2050	-0.1180
WOB			1.0000	-0.1670	0.0765	-0.0889	0.6010	0.1914	0.5779	0.1278	-0.0790	-0.0673	0.0561	0.4535	0.4261	0.0909	0.5794	-0.1228
ROT				1.0000	-0.0214	0.3818	-0.3719	-0.1116	-0.3366	0.1126	0.1083	0.1914	0.0921	-0.0997	-0.0790	0.1495	-0.4333	-0.0342
TOQ					1.0000	0.0887	0.2590	0.0584	0.2504	0.0616	-0.0238	0.0597	0.1573	0.0509	0.0879	0.0590	0.1676	0.1378
SP						1.0000	-0.2203	-0.0715	-0.1762	0.3940	0.3313	0.4756	0.2195	-0.1149	-0.1064	0.4777	-0.4326	0.0396
Fin							1.0000	0.2233	0.9771	0.0373	-0.2790	-0.2483	-0.1179	0.3757	0.3511	0.0018	0.8482	-0.1181
Fout								1.0000	0.2421	0.0660	0.0841	0.0117	0.0286	0.0984	0.0951	0.0375	0.1343	-0.0540
STK									1.0000	0.0465	-0.2859	-0.2335	-0.1110	0.3535	0.3320	0.0156	0.7819	-0.1036
MW										1.0000	0.5536	0.8506	0.2937	0.0527	0.0461	0.9475	-0.0531	-0.1177
Fvis											1.0000	0.7357	0.6201	0.1015	0.0660	0.5397	-0.2615	-0.0560
Pvis												1.0000	0.4956	-0.0349	-0.0361	0.8198	-0.3005	0.0250
YP													1.0000	0.4441	0.4378	0.2564	-0.0901	0.0514
GS10s														1.0000	0.9174	0.0003	0.4871	-0.1601
GS10m															1.0000	-0.0017	0.4640	-0.1481
SLD																1.0000	-0.1051	-0.0715
HD																	1.0000	-0.1236
Loss																		1.0000

Fig. 4 Correlation coefficient matrix for the variables considered in the Azadegan oil field lost circulation dataset**Table 3** Description of the twelve ML/DL models applied to classify the Azadegan field loss circulation dataset

Algorithm name	Abbreviation	Description	Lost circulation studies applying algorithm
Adaboost	ADA	Boosted decision tree	Magzoub et al. (2021)
Convolutional neural network	CNN	Deep learning neural network	Wang et al. (2020); Aljubran et al. (2021); Mardanirad et al. (2021)
Decision tree	DT	A single tree configuration	Sabah et al. (2019); Alkinani et al. (2020b, a)
Gated recurrent unit	GRU	Deep learning neural network	Mardanirad et al. (2021)
K-nearest neighbour	KNN	Data matching algorithm	Cristofaro et al. (2017); Magzoub et al. (2021)
Long short-term memory	LSTM	Deep learning neural network	Aljubran et al. (2021); Mardanirad et al. (2021)
Multi-layer perceptron	MLP	Classic neural network	Cristofaro et al. (2017); Hou et al. (2020); Sabah et al. (2021)
Naïve Bayes classifier	NBC	Simple Bayes' theorem classifier	Liang et al. (2014); Cristofaro et al. (2017)
Quadratic discriminant analysis	QDA	Nonlinear classifier	Dankowska and Kowalewski (2019) provide an example of its application, but it has not been previously applied to lost circulation
Random forest	RF	Tree-based ensemble method using bagging	Li et al. (2018); Aljubran et al. (2021); Magzoub et al. (2021)
Support vector classifier	SVC	Non-probabilistic classifier	Behnoudfar and Hosseini (2016); Manshad et al. (2017); Ahmed et al. (2019)

presented here. Instead, readers are referred to published sources that provide the detailed methodology for each algorithm evaluated.

Each DL/ML algorithm considered requires specific configurations. These mainly constitute tuning of their control variables to modify their architecture and/or learning rates to work effectively with compiled dataset. The most effective control parameter settings have been obtained by multiple trial-and-error evaluations of each algorithm.

The DL/ML methods evaluated in this study (except QDA) have been previously applied, in various configurations, to predict lost circulation datasets in several published studies, which are cited in Table 3. Those cited studies provide methodology details and mathematical descriptions for each method. The majority of the studies cited apply either one method or compare a small selection of ML/DL methods. One novelty of this study is that compares the performance of eleven different ML/DL methods in the prediction of fluid-loss classes applied to the same large field dataset. For this study, the twelve models considered are configured with the control parameter values described in Table 4.

Class prediction accuracy measures assessed

To access the loss classification prediction performance of the ML/DL models evaluated five distinctive statistical measures of accuracy commonly assessed for classification-type problems are used in this study. These are:

- The absolute number of prediction errors recorded
- *Accuracy* (A) as expressed in Eq. (1):

$$A = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

where, in terms of numbers of predictions, TP = true positive predictions, TN = true negative predictions, FP = false positive predictions and FN = false negative predictions.

- *Precision* (P) as expressed in Eq. (2):

$$P = TP/(TP + FP) \quad (2)$$

- *Recall* (R) as expressed in Eq. (3):

$$R = TP/(TP + FN) \quad (3)$$

Table 4 Control parameters applied to twelve ML/DL models configured to predict fluid-loss classes in the Azadegan field dataset

Machine/deep learning algorithms	Control parameter values applied	Model	Software
Adaboost (ADA)	Number of estimators = 2000; learning rate = 0.2; base estimator is DT with depth = 1000; splitter = best	Scikit-learn	Python
Convolutional neural network (CNN1)	1D Convolutional layer = 5 (filters = 200; size = 3; activation fn. = relu) Dropout = 0.5 and Flatten (between 1D and Dense layers) Dense Layers = 2 (neurons = 500; activation = relu) Optimizer = adam (learning rate = 0.001); loss function = mse Iterations = 100; batch size = 32	Keras/TensorFlow	Python
Convolutional neural network (CNN2)	1D Convolutional layer = 1 (filters = 64; size = 3; activation fn. = relu) MaxPooling Layer = 1 (pool size = 1) Dropout = 0.5 and Flatten (between 1D and Dense layers) Dense Layers = 2 (neurons = 100; activation = relu) Optimizer = adam (learning rate = 0.001); loss function = mse Iterations = 250; batch size = 32	Keras / TensorFlow	Python
Decision tree (DT)	Maximum depth = 100; splitter = best; loss function = entropy	Scikit-learn	Python
Gated recurrent units (GRU)	GRU Layer = 1: hidden nodes = 32; dropout = 0.6; recurrent dropout = 0.2 Dense layers = 2: 50 nodes; activation fn. = relu Output (Dense) Layer: neurons = 1 Optimizer = adam (learning rate = 0.001); loss function = mse Iterations = 250; batch size = 32	Keras/TensorFlow	Python
K-nearest neighbour (KNN)	Neighbours = 10; weighted by Euclidian distance	Scikit-learn	Python
Long short-term memory (LSTM)	LSTM Layer = 1: hidden nodes = 32; dropout = 0.5 Dense layers = 2: 50 nodes; activation fn. = relu Output (dense) Layer: neurons = 1 Optimizer = adam (learning rate = 0.001); loss function = mse Iterations = 250; batch size = 32	Keras/TensorFlow	Python
Multi-layer perceptron (MLP)	3 hidden layers with 200, 100 and 50 neurons; activation fn. = logistic; Solver = adam; alpha = 0.0001; maximum iterations 1000; adaptive learning rate	Scikit-learn	Python
Naïve Bayes classifier (NBC)	Variable smoothing = 1e-9	Scikit-learn	Python
Quadratic discriminant analysis (QDA)	No control parameters involved		
Random forest (RF)	Number of estimators = 250; Maximum depth = 100	Scikit-learn	Python
Support vector regression (SVR)	Kernel = rbf; C = 5000; gamma = 0.00001; epsilon = 0.01	Scikit-learn	Python

- *Balanced F-score* (F_1), the harmonic mean of P and R , as expressed in Eq. (4):

$$F_1 = (2 \times (R \times P)) / (R + P) \quad (4)$$

In addition to these five commonly used measures of classification accuracy, confusion matrices, customized for this study, offer a useful visual dimension with which to simultaneously view the prediction accuracy and errors associated with each of the five fluid-loss classes. The top-left to bottom-right diagonal of the confusion matrix displays the TP predictions. In a perfect prediction performance all predictions are located along that diagonal of the confusion matrix.

Four other prediction performance accuracy metrics are also used to assess and compare the training and testing performance of each ML/DL model evaluated. These are:

- *Mean Square Error* (MSE) as expressed in Eq. (5):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2 \quad (5)$$

where X_i = actual recorded loss class value of data record i , and Y_i = predicted loss class value of data record i . MSE is used as the objective function in the training process of the ML/DL models.

- *Root Mean Square Error* (RMSE) as expressed in Eq. (6):

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (6)$$

- *Correlation Coefficient* (R) between actual X_i and predicted Y_i values (on a scale between -1 and $+1$) as expressed in Eq. (7):

$$R = \frac{\sum_{i=1}^n (X_i - X_{\text{mean}})(Y_i - Y_{\text{mean}})}{\sqrt{\sum_{i=1}^n (X_i - X_{\text{mean}})^2 \sum_{i=1}^n (Y_i - Y_{\text{mean}})^2}} \quad (7)$$

- *Coefficient of Determination* (R^2) (on scale between 0 and 1) as expressed as the value of R from Eq. (7) squared.

Pre-processing and process workflow applied to the fluid-loss dataset

The original dataset (Mardanirad et al. 2021) included 22 variables and several incomplete data records. The data records were filtered to remove incomplete data records and omit descriptive variables using qualitative string (text) descriptions. Also, data records associated with mechanical equipment problems (e.g. mud-pump failures) were removed. After data filtering, the dataset was reduced to

65,376 data records (Tables 1, 2 and 3) involving 17 input variables with complete and valid data values. The dependent variable “loss” was transformed from string descriptions of the five fluid-loss classes into numerical (1–5) values for each data record. That numerical transformation made it possible to use the mean squared error (MSE) loss function for algorithm optimization.

The filtered data record variables were then normalized to a range of -1 to $+1$, thereby avoiding scaling biases impacting the algorithms. Normalization was achieved using Eq. (8).

$$\text{Norm}x_i^m = 2 * \left(\frac{x_i^m - x_{\min}^m}{x_{\max}^m - x_{\min}^m} \right) - 1 \quad (8)$$

where $\text{Norm}x_i^m$ is the normalized variable value, x_i^m is the actual recorded value of the m th variable of i th data record, $x_{\min}^m$ is the minimum value of the m th variable considering all filtered data records and $x_{\max}^m$ is the maximum value of the m th variable considering all filtered data records.

Trial-and-error tests were also conducted with various random splits of the data records between training and testing subsets. These tests established that a random split of 75% of the data records to the training subset and 25% to the testing subset generated meaningful fluid-loss classifications. As the distribution of the fluid-loss classes in the dataset is quite imbalanced, with relatively few data records falling into the complete-loss class, the testing subset needs to be quite large in order to contain a reasonable number of data records falling into that class. The testing subset was held independently, and in no way involved in algorithm training. The trained models are then applied to predict the fluid-loss class of each data record in the testing subset. The dataset was randomly split into training and testing subsets ten times and evaluated by each algorithm and an average RMSE, MAE and R^2 values of the ten random cases taken as representative of each model’s prediction performance.

Figure 5 illustrates diagrammatically the work flow applied in executing and evaluating the ML/DL algorithms to predict fluid-loss classes for the Azadegan oil field dataset.

Results

Table 5 lists the loss class prediction accuracy achieved by the 12 ML/DL models evaluated for the training and testing subsets and the complete dataset in terms of RMSE, R^2 and absolute number of prediction errors (#Errors). The model execution times are also listed in Table 5. It is clear that the models vary substantially in terms of both their prediction accuracy and execution speeds.

The tree-based models (RF, DT and ADA) and data matching (KNN) ML models outperform the correlation/

Fig. 5 Lost circulation workflow for ML/DL models

Table 5 Prediction accuracy comparison of the twelve ML/DL models evaluated to predict fluid-loss classes in the Azadegan field dataset

Training and testing prediction accuracy of twelve algorithms compared for the Azadegan reservoir lost circulation dataset										
Number of data records	Training subset (75%)			Testing subset (25%)			All data			Execution
	49,032			16,344			65,376			
ML/DL Algorithm	R^2	RMSE	#Errors	R^2	RMSE	#Errors	R^2	RMSE	#Errors	Sec
Random forest	1	0	0	0.9850	0.0686	35	0.9963	0.0343	35	36.29
Decision tree	1	0	0	0.9722	0.0935	73	0.9931	0.0468	73	3.88
Adaboost	1	0	0	0.9725	0.0929	84	0.9932	0.0464	84	4.06
K-nearest neighbour	1	0	0	0.9714	0.0948	99	0.9929	0.0474	99	62.49
Support vector classifier	0.9968	0.0319	44	0.9573	0.1158	124	0.9871	0.0641	168	70.72
Convolutional neural network (1)	0.9796	0.0814	166	0.9631	0.1066	359	0.9756	0.0968	525	7797.96
Convolutional neural network (2)	0.9433	0.1355	959	0.9386	0.1367	329	0.9422	0.1475	1288	1225.18
Gated recurrent unit	0.9481	0.1281	816	0.9373	0.1435	341	0.9545	0.1433	1157	3422.53
Long short-term memory	0.9206	0.1604	1242	0.9141	0.1614	446	0.9192	0.1746	1688	3286.06
Multi-layer perceptron	0.4743	0.4105	6065	0.4576	0.4128	2049	0.4702	0.4110	8114	178.69
Quadratic discriminant analysis	0.0475	0.5525	13,129	0.0301	0.5520	4381	0.0433	0.5524	17,510	3.90
Naïve Bayes classifier	−0.854	0.7708	20,974	−0.9680	0.7863	7108	−0.882	0.7747	28,082	3.57

The algorithms are listed in order of their prediction performance in relation to the testing subset: the model with the lowest #Errors listed first

regression-based models in terms of all measures of fluid-loss class prediction accuracy (Tables 5 and 6). SVR is the best performing regression-based model, comfortably outperforming the DL models. The more complex CNN1 model provides the best prediction accuracy of the DL models, which all achieve testing subset #Errors < 500 and $R^2 > 0.9$. The three poorest performing models (MLP, QDA and NBC) do not result in good prediction outcomes for this dataset with #Errors > 2000 and $R^2 < 0.5$ for the testing subset.

The model execution times, on the same laptop (1.8 GHz microprocessor with 4 MB cache and 2 cores), expressed in seconds, vary from 3.57 (NBC) to 7798 (CNN1). The complexity of the DL models is reflected in their execution

times. The more complex architecture of the CNN1 model results in an execution time that is more than six times longer than the CNN2 model. Indeed, there is more than two order of magnitude difference in the execution speeds between the faster and the slowest models. Whereas the simpler regression-based models (QDA and NBC) execute in 3–4 s, they do not deliver useful fluid-loss class prediction accuracy. On the other hand, two of the tree-based models (ADA and DT) also execute in 3–4 s and those models do deliver high prediction accuracy (Table 5).

The RF model delivers the best prediction accuracy in terms of #Errors for the testing subset (just 35) and executes in 36.29 s (Table 5). The top-five models in terms

Table 6 displays the prediction performance of the six most accurate models in terms of their A, P, R and F1 metrics for all data records and representative testing subsets

Accuracy and precision achieved by six best performing models compared for the Azadegan reservoir lost circulation dataset										
Class	Full dataset					Testing subset				
	65,376 data records					16,344 data records				
	1	2	3	4	5	1	2	3	4	5
<i>Random forest</i>										
Accuracy	0.9997	0.9995	0.9998	1.0000	1.0000	0.9989	0.9982	0.9990	0.9998	0.9998
Precision	0.9997	0.9994	0.9962	1.0000	0.9697	0.9987	0.9975	0.9855	1.0000	0.6667
Recall	1.0000	0.9983	0.9977	0.9889	0.9412	0.9998	0.9930	0.9913	0.9492	0.5000
F_1 score	0.9998	0.9988	0.9970	0.9944	0.9552	0.9993	0.9952	0.9884	0.9739	0.5714
<i>Decision tree</i>										
Accuracy	0.9993	0.9991	0.9995	0.9999	1.0000	0.9971	0.9963	0.9980	0.9997	0.9999
Precision	0.9994	0.9981	0.9936	0.9926	1.0000	0.9977	0.9921	0.9753	0.9655	1.0000
Recall	0.9996	0.9973	0.9943	0.9889	0.9412	0.9986	0.9889	0.9781	0.9492	0.5000
F_1 score	0.9995	0.9977	0.9940	0.9907	0.9697	0.9981	0.9905	0.9767	0.9573	0.6667
<i>Adaboost</i>										
Accuracy	0.9990	0.9989	0.9997	1.0000	0.9999	0.9961	0.9954	0.9986	0.9999	0.9997
Precision	0.9992	0.9975	0.9966	1.0000	0.9143	0.9968	0.9899	0.9868	1.0000	0.4000
Recall	0.9995	0.9967	0.9951	0.9926	0.9412	0.9981	0.9864	0.9810	0.9661	0.5000
F_1 score	0.9994	0.9971	0.9958	0.9963	0.9275	0.9974	0.9881	0.9839	0.9828	0.4444
<i>K-nearest neighbour</i>										
Accuracy	0.9991	0.9987	0.9993	0.9999	1.0000	0.9966	0.9946	0.9973	0.9996	0.9998
Precision	0.9993	0.9977	0.9891	0.9852	0.9697	0.9971	0.9907	0.9586	0.9333	0.6667
Recall	0.9996	0.9954	0.9943	0.9889	0.9412	0.9984	0.9814	0.9781	0.9492	0.5000
F_1 score	0.9994	0.9966	0.9917	0.9871	0.9552	0.9977	0.9860	0.9683	0.9412	0.5714
<i>Support vector classifier</i>										
Accuracy	0.9982	0.9978	0.9990	0.9999	1.0000	0.9942	0.9937	0.9973	0.9998	0.9998
Precision	0.9988	0.9943	0.9875	0.9926	0.9697	0.9963	0.9830	0.9679	0.9825	0.6667
Recall	0.9988	0.9946	0.9868	0.9889	0.9412	0.9961	0.9845	0.9679	0.9492	0.5000
F_1 score	0.9988	0.9944	0.9872	0.9907	0.9552	0.9962	0.9837	0.9679	0.9655	0.5714
<i>Convolutional neural network (1)</i>										
Accuracy	0.9960	0.9927	0.9959	0.9995	0.9999	0.9941	0.9911	0.9953	0.9993	0.9999
Precision	0.9971	0.9829	0.9487	0.9094	1.0000	0.9958	0.9762	0.9489	0.8906	1.0000
Recall	0.9976	0.9800	0.9509	0.9667	0.7353	0.9965	0.9780	0.9275	0.9194	0.7500
F_1 score	0.9974	0.9814	0.9498	0.9372	0.8475	0.9962	0.9771	0.9381	0.9048	0.8571

The models are listed in order of performance (best first) based on their testing subset #Errors

of prediction accuracy all achieve perfect or near-perfect prediction accuracy for the training subset. This is indicative that these models are over-fitting the data to some extent. Despite that evidence of over-fitting these models consistent go on, in multiple random splits of the complete dataset, to predict the testing subsets (held independent of the training process) with substantially lower errors than the other models. Comparison of the training and testing subset results for the other models indicates that they have not over-fitted the training subset, yet deliver poorer prediction accuracy for the testing subset than the top-five performing models.

Consideration of the prediction accuracies in terms of the A , P , R and F_1 metrics (Table 6) reveals some key differences between the performances of the models in terms of predicting the specific fluid-loss classes. The RF model clearly outperforms all other models in terms of its prediction performance for fluid-loss classes 1, 2 and 3 (i.e. the classes where most the data records reside) However, in terms of the P and F_1 metrics the DT model achieves the best prediction performance for the complete fluid-loss class 5, and the ADA model prediction performance for the severe fluid-loss class 4 outperforming the other models. Although in statistical terms, due to the much smaller numbers of data records they relate to, these two classes are less significant to the overall prediction performance of the models, in decision-making terms. Classes 4 and 5 are the two categories of fluid losses that need to be predicted quickly and reliably so that mitigation actions can be promptly undertaken to resolve them. These results suggest that, from an application perspective, there is potential value in executing the RF, DT and ADA models in an ensemble comparing their predictions to provide best prediction accuracy for all five fluid-loss classes.

Discussion

The results presented in Tables 5 and 6 clearly reveal that the tree-based ML models outperform the other ML/DL model evaluated for fluid-loss classification predicted from drilling variables associated with the large and imbalanced Azadegan oil field dataset. Although the basic prediction accuracy measures (RMSE, R^2 and #Errors) are useful for distinguishing the models delivering overall high prediction accuracy (Table 5), it is essential to consider the more specific classification accuracy metrics (A , P , R and F_1) to reveal the models that deliver the best prediction accuracy (RF, DT and ADA) for each fluid-loss class (Table 6). However, in order to gain a better understanding of the detailed prediction performance of each model evaluated confusion matrices need to be considered.

For this study, a novel, expanded confusion matrix format is developed (Figs. 6 and 7) that combines the basic

confusion matrix with details of error numbers and precision achieved for each fluid-loss class. This format, together with the class accuracy metric values (Table 6) makes it easier to visualize each model's performance in more detail when considered for the full dataset or testing subset predictions.

Figure 6 displays the expanded confusion matrices for the best-six performing models, as ranked in Table 5. They more clearly display the class imbalance of the dataset and the superior performance of the RF model, particularly in terms of the most abundant fluid-loss classes 1, 2 and 3. For class 4, the RF and ADA models both achieve 100% precision (i.e. no false positives) but ADA achieves one less false negative and therefore a slightly higher recall (and F_1 score) than the RF model. For class 5, the DT model is the only one to achieve 100% precision (i.e. no false positives). However, the top-5 performing models all correctly identify 32 of the 34 data records in class 5 as belonging to that class (i.e. they all achieve a recall of 0.9412, Table 6), and all confuse their two erroneous data record predictions as belonging to class 1.

Whereas the top-five performing models all achieve high prediction performances for fluid-loss classes 4 and 5 (267 or 268 correct predictions for class 4; 32 out of 34 correct predictions for class 5), model CNN1 performs less well in its predictions for these two classes (261 correct predictions for class 4; 25 correct predictions for class 5). Also, the CNN1 model involves substantially more false positives for class 4 resulting in its lower precision score of 0.9094 for that class. Although the CNN1 model achieves a perfect precision score for class 5 (no false positives), its predictions involve nine false negatives for that class resulting in much lower R (0.7353) and F_1 (0.8475) scores than the other models shown in Fig. 6.

The poorer prediction performance of classes 4 and 5 data records is a characteristic of all the DL models as revealed by the expanded confusion matrices for those models compared in Fig. 7. Indeed, the GRU and LSTM models fail to predict any of the class 5 data records. The GRU model performs little better for class 4 with only 1 correct prediction out of 270 class 4 data records. The LSTM model performs slightly better for class 4 with 99 correct predictions for that class, but this is also substantially lower than the high-performing models. The two CNN models provide more accurate predictions than the GRU and LSTM models, making them the best performing DL models considered. The simpler CNN2 model actually achieves better prediction results for fluid-loss class 5 than the CNN1 model (29 versus 25 correct predictions out of 34). However, the more complex CNN2 model outperforms the CNN1 model in its predictions for all other classes.

It is apparent that despite their network structural complexities and capabilities to handle large datasets, the DL models struggle to adequately process this highly

Full Dataset Confusion Matrices for Six Best Performing Models Compared for the Azadegan Reservoir Lost Circulation Dataset													
65376	RF	49559	12866	2651	267	33	65376	DT	49556	12870	2649	269	32
Total	Class	1	2	3	4	5	Total	Class	1	2	3	4	5
49545	1	49543	2	0	0	0	49545	1	49527	15	2	1	0
12880	2	13	12858	9	0	0	12880	2	21	12845	14	0	0
2647	3	1	4	2641	0	1	2647	3	5	9	2632	1	0
270	4	0	2	1	267	0	270	4	1	1	1	267	0
34	5	2	0	0	0	32	34	5	2	0	0	0	32
Precision: 99.97% 99.94% 99.62% 100.00% 96.97%							Precision: 99.94% 99.81% 99.36% 99.26% 100.00%						
35	#Errors	16	8	10	0	1	73	Errors	29	25	17	2	0
99.95% % Successful Predictions							99.89% % Successful Predictions						
65376	ADA	49561	12869	2643	268	35	65376	KNN	49561	12850	2661	271	33
Total	Class	1	2	3	4	5	Total	Class	1	2	3	4	5
49545	1	49521	21	3	0	0	49545	1	49525	20	0	0	0
12880	2	37	12837	6	0	0	12880	2	31	12821	28	0	0
2647	3	1	9	2634	0	3	2647	3	3	7	2632	4	1
270	4	0	2	0	268	0	270	4	0	2	1	267	0
34	5	2	0	0	0	32	34	5	2	0	0	0	32
Precision: 99.92% 99.75% 99.66% 100.00% 91.43%							Precision: 99.93% 99.77% 98.91% 98.52% 96.97%						
84	Errors	40	32	9	0	3	99	Errors	36	29	29	4	1
99.87% % Successful Predictions							99.85% % Successful Predictions						
65376	SVC	49545	12884	2645	269	33	65376	CNN1	49569	12842	2653	287	25
Total	Class	1	2	3	4	5	Total	Class	1	2	3	4	5
49545	1	49487	51	6	1	0	49545	1	49426	118	1	0	0
12880	2	44	12810	26	0	0	12880	2	131	12622	127	0	0
2647	3	12	21	2612	1	1	2647	3	10	97	2517	23	0
270	4	0	2	1	267	0	270	4	1	2	6	261	0
34	5	2	0	0	0	32	34	5	1	3	2	3	25
Precision: 99.88% 99.43% 98.75% 99.26% 96.97%							Precision: 99.71% 98.29% 94.87% 90.94% 100.00%						
168	Errors	58	74	33	2	1	525	Errors	143	220	136	26	0
99.74% % Successful Predictions							99.20% % Successful Predictions						

Fig. 6 Expanded confusion matrices for the six best performing models in classifying fluid-loss classes for the Azadegan oil field dataset

imbalanced classification task with the fluid-loss dataset evaluated. They seem to focus their processing powers on the most populated classes (classes 1, 2 and 3) at the expense of the sparsely populated classes (classes 4 and 5). Consequently they must be considered as unreliable in terms of correctly predicting severe and complete fluid-loss incident from the drilling variables available for the dataset evaluated. The high model execution times, particularly for the CNN1 model, also pose a problem. In a real-time situation, while drilling, it is necessary to identify fluid-loss classes rapidly in order to make prompt mitigation decisions; in seconds (as achieved by RF, DT and ADA models) not by an hour or more as required by all but one of the DL models.

The technique presented for the Azadegan oil field dataset, and the novel expanded confusion matrix developed, could be applied more generally to other partially developed oil and gas fields to predict fluid-loss classes from large datasets of drilling variables and assess classification prediction performance in detail. Clearly, the datasets would require careful training and testing and several models should

ideally be evaluated to identify the best performing model for each fluid-loss class. The results from this study suggest that it is incorrect to assume that large datasets, particularly those with highly imbalanced class distributions will always be better predicted by DL methods compared to ML methods. Hence, ML methods, particularly those that are tree-based or involve data matching should not be ruled out on the grounds of dataset size considerations only. Moreover, when comparing DL model performance against ML performance for classification problems, that comparison should include tree-based and data matching methods as well as neural networks and other regression-based models.

Conclusions

The prompt and accurate prediction of wellbore fluid losses based on drilling variables is desirable as it can help to reduce drilling costs, time and mitigate fluid-loss events from becoming serious well-control problems. Drilling

Full Dataset Confusion Matrices for Four Deep Learning Models Compared for the Azadegan Reservoir Lost Circulation Dataset													
65376	CNN1	49569	12842	2653	287	25	65376	CNN2	49719	12726	2609	290	32
Total	Class	1	2	3	4	5	Total	Class	1	2	3	4	5
49545	1	49426	118	1	0	0	49545	1	49243	299	3	0	0
12880	2	131	12622	127	0	0	12880	2	462	12190	228	0	0
2647	3	10	97	2517	23	0	2647	3	10	235	2369	33	0
270	4	1	2	6	261	0	270	4	0	2	8	257	3
34	5	1	3	2	3	25	34	5	4	0	1	0	29
Precision: 99.71% 98.29% 94.87% 90.94% 100.00%							Precision: 99.04% 95.79% 90.80% 88.62% 90.63%						
525	Errors	143	220	136	26	0	1288	Errors	476	536	240	33	3
99.20% % Successful Predictions							98.03% % Successful Predictions						
65376	GRU	49614	13109	2629	24	0	65370	LSTM	49936	13231	2070	131	2
Total	Class	1	2	3	4	5	Total	Class	1	2	3	4	5
49545	1	49343	197	5	0	0	49545	1	49439	105	1	0	0
12880	2	256	12572	52	0	0	12880	2	465	12330	85	0	0
2647	3	12	332	2303	0	0	2638	3	27	792	1814	5	0
270	4	1	6	262	1	0	272	4	1	3	167	99	2
34	5	2	2	7	23	0	35	5	4	1	3	27	0
Precision: 99.45% 95.90% 87.60% 4.17% 0.00%							Precision: 99.00% 93.19% 87.63% 75.57% 0.00%						
1157	Errors	271	537	326	23	0	1688	Errors	497	901	256	32	2
98.23% % Successful Predictions							97.42% % Successful Predictions						

Fig. 7 Expanded confusion matrices for the four DL models in classifying fluid-loss classes for the Azadegan oil field dataset

datasets in relation to wellbore fluid-loss classes suffer from two significant characteristics: 1) individual drilling variables tend not to be highly correlated with fluid-loss classes identified, and 2) the distribution of data records across fluid-loss classes tend to be highly imbalanced with the majority located in the no/seepage-loss classes and very few in the severe/complete-loss classes. The first characteristic limits the class prediction accuracy that can be achieved by many regression-based models. The second characteristic means that prediction accuracy of models assessed purely in terms of total prediction errors can be misleading as it may mask poor prediction accuracy of the more sparsely represented classes in a dataset.

This study uses the large, 20-well dataset (65,376 data records) for the Azadegan oil field (Iran) with 17 drilling variables and 5 wellbore fluid-loss classes distinguished, to assess the prediction performance of twelve models related to eight machine learning (ML) and three deep learning (DL) techniques. Of these models, it is the tree-based algorithms (random forest (RF), decision tree (DT), Adaboost (ADA)) and K-nearest neighbour (KNN) data matching algorithms that deliver the best predictions (i.e. < 100 errors for all data records and only 35 errors for the RF model). These models outperform the other regression-based ML/DL models in their prediction accuracy and require much shorter execution times than the DL models.

Considering class accuracy metrics precision, recall and F1 score reveal that for wellbore fluid-loss classes 1 to 3 (less significant losses and most data records) the RF

model is most accurate. On the other hand, for wellbore fluid-loss classes 4 and 5 (more significant losses and fewest data records) the ADA and DT models, respectively, are the most accurate. These results imply that it would make sense in practice to run all three of these tree-based models as an ensemble to provide prediction comparisons for each fluid-loss class.

Model class prediction accuracy can be usefully displayed using an expanded confusion matrix format to integrate and visualize the models performance in terms of accuracy and their ability to address the class imbalance characteristic of lost circulation datasets. These displays reveal the superior performance of the ML models mentioned to the DL models. In particular, the DL models (CNN, GRU and LSTM) are substantially inferior in their prediction accuracy for fluid-loss classes 4 and 5. The CNN model can be improved to an extent by adding layers but that adversely increases its complexity and execution time, and that modified model is still unable to match the performance of the high-performing ML methods with the Azadegan dataset.

The key findings of this study are summarized as:

- An ensemble of tree-based ML models (RF, ADA and DT) provides the best prediction accuracy of loss circulation classes;
- That ensemble overcomes the class imbalance issues typically posed by loss circulation datasets;
- ML algorithms are more accurate and quicker to exercise than DL algorithms for the large dataset evaluated;

- DL algorithms perform particularly poorly for those classes with the fewest data records; and
- The novel expanded confusion matrix display provides excellent visualization of classification accuracy and class imbalance performance for loss circulation datasets.

Funding No funding was received for this study.

Data and material availability The dataset evaluated is published (Mardanirad et al. 2021).

Code availability Publicly available Python code functions were used in the models evaluated and control parameters applied to those are specified in the manuscript.

Declarations

Conflicts of interest The authors have no conflicts of interest associated with this 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/>.

References

- Abel LW, Gokhan Y (2004) Underbalanced techniques to reduce drilling hazards of loss of circulation and underground flows. In: SPE annual technical conference and exhibition, Houston, Texas, 2004. Doi: <https://doi.org/10.2118/89507-MS>
- Agin F, Khosravanian R, Karimifard M, Jahanshahi A (2020) Application of adaptive neuro-fuzzy inference system and data mining approach to predict lost circulation using DOE technique (case study: Maroon oil field, Iran). *Petroleum* 6(4):423–437. <https://doi.org/10.1016/j.petlm.2018.07.005>
- Ahmed A, Elkatatny S, Abdulraheem A, Abughaban M (2019) Prediction of lost circulation zones using support vector machine and radial basis function. In: International petroleum technology conference, Dhahran, Kingdom of Saudi Arabia. Doi: <https://doi.org/10.2523/IPTC-19628-MS>
- 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
- 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 Xplore Access* 9:76833–76846. <https://doi.org/10.1109/ACCESS.2021.3082557>
- Alkinani H, Alhameedi A, Dunn-Norman S, Flori RE, Alsaba MT, Amer AS, Hilgedick SA (2018) Using data mining to stop or mitigate lost circulation. *J Pet Sci Eng*. <https://doi.org/10.1016/j.petrol.2018.10.078>
- Alkinani H, Alhameedi A, Dunn-Norman S, Alkhamis M, Mutar R (2019a) Prediction of lost circulation prior to drilling for induced fractures formations using artificial neural networks. *SPE*. SPE-195197-MS.
- Alkinani H, Alhameedi T, Dunn-Norman S, Alalwani M, Mutar RA, AlBazzaz WH (2019b) Data-driven neural network model to predict equivalent circulation density (ECD). *SPE*, SPE-198612-MS.
- Alkinani H, Al-Hameedi ATT, Dunn-Norman S, Al-Bazzaz WH (2020) State-of-the-art review of lost circulation materials and treatments—part II: probability and cost analyses. In: International Petroleum technology conference, Dhahran, Saudi Arabia, 2020. Doi: <https://doi.org/10.2523/IPTC-19877-MS>
- Alkinani H, Alhameedi T, Dun-Norman S (2020b) Data-driven decision-making for lost circulation treatments: a machine learning approach. *J Energy AI*. <https://doi.org/10.1016/j.egyai.2020.100031>
- 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* 26:627–634. <https://doi.org/10.1016/j.ejpe.2016.09.004>
- Choubineh A, Wood DA, Choubineh Z (2020) Applying separately cost-sensitive learning and fisher's discriminant analysis to address the class imbalance problem: a case study involving a virtual gas pipeline SCADA system. *Int J Crit Infrastruct Prot* 29:100357. <https://doi.org/10.1016/j.ijcip.2020.100357>
- 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. In: Offshore technology conference Rio de Janeiro, Brazil. DOI: <https://doi.org/10.4043/28034-MS>
- Dankowska A, Kowalewski W (2019) Comparison of different classification methods for analyzing fluorescence spectra to characterize type and freshness of olive oils. *Eur Food Res Technol* 245:745–752. <https://doi.org/10.1007/s00217-018-3196-z>
- Du Y, Chen J, Cui Y, Xin J, Wang J, Li YZ (2016) Genetic mechanism and development of the unsteady Sarvak play of the Azadegan oil field, southwest of Iran. *Pet Sci* 13:34–51. <https://doi.org/10.1007/s12182-016-0077-6>
- Feng Y, Gray KE (2018) Modeling lost circulation through drilling-induced fractures. *SPE J* 23(01):205–223. <https://doi.org/10.2118/187945-PA>
- Guo X, Yin Y, Dong C, Yang G, Zhou G (2008) On the class imbalance problem. In: Fourth international conference on natural computation pp. 192–201. Doi: <https://doi.org/10.1109/ICNC.2008.871>
- 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. In: Offshore technology conference, Houston, USA. Doi: <https://doi.org/10.4043/30653-MS>
- Jahanbakhshi R, Keshavarzi R (2015) 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>
- Lavrov A (2013) Numerical modeling of steady-state flow of a non-newtonian power-law fluid in a rough-walled fracture. *Comput Geotech* 50:101–109

- SciKit Learn (2021) Supervised and unsupervised machine learning models in Python. <https://scikit-learn.org/stable/> [visited 20th September 2021]
- 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, Seattle, Washington, USA, ARMA-2018-105
- 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
- Ling CX, Sheng VS (2011) Cost-sensitive learning and the class imbalance problem. In: Sammut C, Webb GI (eds) *Encyclopedia of machine learning*. Springer, Boston. <https://doi.org/10.1007/978-0-387-30164-8>
- Liu H, Guo R, Dong J, Liu L, Liu Y, Yi Y (2013) Productivity evaluation and influential factor analysis for Sarvak reservoir in South Azadegan oil field. *Iran Petrol Explor Dev* 40(5):627–634
- Magzoub MI, Kiran R, Salehi S, Hussein IA, Nasser MS (2021) Assessing the relation between mud components and rheology for loss circulation prevention using polymeric gels: a machine learning approach. *Energies* 14(5):1377. <https://doi.org/10.3390/en14051377>
- Maheshwari S, Jain JC, Jadon RS (2017) A review on class imbalance problem: analysis and potential solutions. *Int J Comput Sci Issues*. <https://doi.org/10.20943/01201706.4351>
- 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>
- Manshad AK, Rostami H, Niknafs H (2017) Integrated lost circulation prediction in oil field drilling operation. In: AH Mohammadi (Ed) Chapter 12 in heavy oil: characteristics, production and emerging technologies, Nova Science Publishers, Inc., NY, USA, pp. 243–253
- Mardanirad S, Wood DA, Zakeri H (2021) The application of deep learning algorithms to classify subsurface drilling lost circulation severity in large oil field datasets. *SN Appl Sci* 3:785. <https://doi.org/10.1007/s42452-021-04769-0>
- 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>
- Pirizadeh M, Alemohammada N, Manthouri M, Pirizadeh M (2021) A new machine learning ensemble model for class imbalance problem of screening enhanced oil recovery methods. *J Petrol Sci Eng* 198:108214. <https://doi.org/10.1016/j.petrol.2020.108214>
- 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* 177:236–249. <https://doi.org/10.1016/j.petrol.2019.02.045>
- Sabah M, Mehrad M, Ashrafi SB, Wood DA, Fathi S (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>
- Salehi S, Nygaard R (2011) Evaluation of new drilling approach for widening operational window: implications for wellbore strengthening. In: SPE production and operations symposium, Oklahoma, USA. Doi: <https://doi.org/10.2118/140753-MS>
- Sanfillipo F, Brignoli M, Santarelli FJ, Bezzola C (1997) Characterization of conductive fractures while drilling. In: SPE European formation damage conference, The Hague, Netherlands. Doi: <https://doi.org/10.2118/38177-MS>
- Savari S, Whitfill DL, Walker J (2016) Lost circulation management in naturally fractured reservoirs. In: SPE/IADC middle east drilling technology conference and exhibition, Abu Dhabi, UAE, 2016. Doi: <https://doi.org/10.2118/178165-MS>
- Shahri MP, Mehrabi M (2012) A new approach in modeling of fracture ballooning in naturally fractured reservoirs. In: SPE Kuwait international petroleum conference and exhibition, Kuwait City, Kuwait. Doi: <https://doi.org/10.2118/163382-MS>
- Su X, Lian Z, Xiong FJH, Wu R, Yuan Y (2019) Lost circulation material for abnormally high temperature and pressure fractured-vuggy carbonate reservoirs in Tazhong block, Tarim Basin. *NW China Pet Explor Dev* 46(1):173–180. [https://doi.org/10.1016/S1876-3804\(19\)30017-5](https://doi.org/10.1016/S1876-3804(19)30017-5)
- TensorFlow (2021) Deep learning models in Python. https://www.tensorflow.org/guide/keras/sequential_model [visited 20th September 2021]
- Van Der Zwaag CH (2006) Benchmarking the formation damage of drilling fluids. *SPE Prod Oper* 21(01):40–50. <https://doi.org/10.2118/86544-PA>
- Wood DA (2018) Transparent open-box learning network provides insight to complex systems and a performance benchmark for more-opaque machine learning algorithms. *Adv Geo-Energy Res* 2(2):148–162. <https://doi.org/10.26804/ager.2018.02.04>
- Zhao X, Qiu Z, Sun B, Liu S, Xing X, Wang M (2019) Formation damage mechanisms associated with drilling and completion fluids for deepwater reservoirs. *J Petrol Sci Eng* 173:112–121. <https://doi.org/10.1016/j.petrol.2018.09.098>

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