# Enhancing fault diagnosis of undesirable events in oil & gas systems: A machine learning approach with new criteria for stability analysis and classification accuracy

**Article** *in* Proceedings of the Institution of Mechanical Engineers Part E Journal of Process Mechanical Engineering · November 2023

DOI: 10.1177/09544089231213778



**5 authors**, including:

Sahraoui Mohammed Amine
University M'Hamed Bougara of Boumerdes


Chemseddine Rahmoune
University M'Hamed Bougara of Boumerdes


Mohamed Zair
University M'Hamed Bougara of Boumerdes


Fawzi Gougam
University M'Hamed Bougara of Boumerdes

Original article

Institution of MECHANICAL ENGINEERS

Journal of PROCESS MECHANICAL ENGINEERING

# Enhancing fault diagnosis of undesirable events in oil & gas systems: A machine learning approach with new criteria for stability analysis and classification accuracy

Proc IMechE Part E: J Process Mechanical Engineering
1–11

DOI: 10.1177/09544089231213778
journals.sagepub.com/home/pie
Sage

**Mohammed Amine Sahraoui[1], Chemseddine Rahmoune[2], Mohamed Zair[2], Fawzi Gougam[2] and Ali Damou[2]**


## Abstract

Petroleum serves as a cornerstone of global energy supply, underpinning economic development. Consequently, the effective detection of faults in oil and gas (O&G) wells is of paramount importance. In response to the limitations observed in prior research, this study presents an innovative fault diagnosis system, rooted in machine learning techniques. Our approach encompasses a comprehensive analysis, incorporating stability assessment via standard deviation (STD), and a meticulous evaluation of accuracy and stability for distinct fault scenarios. By integrating data preprocessing, feature selection methods, and deploying a robust random forest classifier, our model achieves a substantial enhancement in fault classification accuracy and stability. Extensive experimentation substantiates the superiority of our approach, surpassing the performance of previous studies that predominantly emphasized overall accuracy while disregarding stability analysis. Notably, our model attains remarkable accuracies, notably achieving a flawless 100% accuracy for scenario 3 faults. Detailed examination of mean accuracies and STDs further reinforces the precision and consistency of our model's predictive capabilities. Additionally, a qualitative assessment underscores the practical utility and reliability of our model in accurately identifying critical fault types. This research significantly advances fault detection methodologies within the O&G industry, providing valuable insights for decision-making systems in oil well operations.



## Keywords

Fault detection, stability analysis, accuracy, oil & gas undesirable events, random forest, marine predator algorithm

Date received: 12 July 2023; accepted: 5 October 2023


## Introduction

Petroleum is considered one of the most important natural energy sources and is considered to be the leading index of world economic development. Worldwide, demand for oil has grown rapidly as it is used in a variety of applications, such as heating and power generation. The OPEC predicts that in 2024, world demand for oil will increase by 4.15 million barrels per day.[1] Offshore oil wells are typically deep, confined holes that combine a collection of sensors, pneumatic, hydraulic, and mechanical systems to transport the oil to the subsurface.[2] A well in an offshore oil field is a particular application where a borehole is drilled below the seabed to transport the oil to the surface, necessitating a more robust design than other oil wells.[3] Offshore oil well sensors generate huge quantities of multivariate time-series data.[4]

Currently, monitoring survey engineers work manually to analyze the data generated on a schedule basis, a process in which one survey engineer can be assigned to several wells. In order to analyze the data exhaustively, survey engineers have to spend more time and resources on data analysis, instead of concentrating on more critical situations.[5] As a result, critical failures have been bypassed by a significant time delay between problem occurrence and detection, exposing oil drilling and processing to the risk of major failures and losses that can affect industries around the world. Consequently, detecting undesirable and unwanted events in oil and gas (O&G) wells is a key factor in preventing production

[1]Systems Engineering and Telecommunications Laboratory (LIST), University M'hamed BOUGARA, Boumerdes, Algeria
[2]Solid Mechanics and Systems Laboratory (LMSS), University M'hamed Bougara, Boumerdes, Algeria

**Corresponding author:**
Mohammed Amine Sahraoui, Systems Engineering and Telecommunications Laboratory (LIST), University M'hamed BOUGARA Boumerdes, Algeria.
Email: ma.sahraoui@univ-boumerdes.dz

downtime, environmental accidents, and loss of human life, as well as reducing maintenance costs.[6,7]

E. Oort et al.[8] have developed method, called automatic rig-activity detection (ARAD), that allows the massive amount of data generated by well sensors to be analyzed more efficiently and accurately. It operates by monitoring drilling operations through real-time data generated by the platform. The ARAD is unable to identify and detect abnormal conditions, since it classifies abnormal activity as "unknown."[9,10] Instead, such situations must be classified as abnormal to assist engineers in identifying well performance and considering the root causes of these anomalies for any future prevention, in order to reduce the possibility of serious injury, loss of life, economic loss, or environmental pollution. For instance, the catastrophic Macondo disaster in 2010 due to safety equipment failure illustrates the magnitude of potential losses and costs.[11,12] The Deepwater Horizon catastrophic rig resulted in the deaths of 11 workers, and in US history was one of the largest environmental disasters due to massive marine and offshore damage.[13] In terms of maintenance cost, the offshore probe to repair a production line, for example, can exceed US $500,000 per day.[14] Consequently, the use of the latest technologies in this field is crucial to minimizing its serious impact.[15]

Machine learning (ML) has attracted much attention by delivering a number of robust techniques with valuable outcomes in a wide range of applications. The detection of anomalies is one of the most widespread problems addressed by ML methods.[16] At now, many methods have been introducing de detect the failure in oil & gas field as industrial 4.0 by integrated the cloud computing, the real-time data analysis, internet of thing, etc. The enormous data set collected from a variety of sensors in many fields has been analyzed and successfully used for prediction in different areas using ML. In the O&G industry, ML techniques have also shown significant results.[17]

This study proposes a condition-based monitoring (CBM) system for O&G wells that assist to decision-making systems. The proposed system is designed to detect and recognize abnormal behaviors at the earliest possible stage, to avoid the production losses and maintenance costs. The study focused on a database developed by Petrobras comprising around 2000 events. In this context, widely researchers have been challenged task to develop method based on MLs for early detection and classification for O&G wells, for instance, Ricardo Emanuel Vaz Vargas developed method to Detect of undesirable events in O&G wells,[6] this contribution based on only detection of undesirable events, which occur over time. In the study,[18] researchers conducted a comparative analysis of two distinct feature extraction methods while employing multiple classifiers in an attempt to augment the outcomes. Conversely, in another investigation mentioned in,[19] researchers focused on enhancing the performance of one-class classifiers when applied to the task of anomaly detection in oil wells.

In contrast to previous research studies focused on the classification of unwanted faults in oil wells,[5,20] it is worth noting that a limited number of these studies have extensively examined the stability and accuracy of the results and the evaluation of performance measures for each class. Additionally, the consideration of standard deviation (STD) as a measure of stability and the incorporation of accuracy and stability analysis for each class has been largely overlooked.

For instance, in the study conducted by Vargas et al.,[6] the focus was solely on the detection of undesirable events over time, without explicit evaluation of stability measures and accuracy for individual classes. Similarly, N. Aslam et al.[21] developed a CBM system achieving an overall accuracy of 90%, but the stability analysis and evaluation of performance measures for each class were not extensively explored. Another research effort by Marins et al.[5] demonstrated an overall accuracy of 94% in detecting and classifying faulty events but did not include a thorough investigation of stability and accuracy for each class. More recently, Castro et al.[22] introduced unsupervised methods for classifying real offshore well data, but the stability analysis using STD and accuracy evaluation for individual classes were not explicitly addressed.

In the context of this research paper, our primary objective is to address certain existing disparities by introducing a comprehensive methodology. This methodology encompasses a thorough investigation of stability, measured through the STD, and a meticulous evaluation of accuracy and stability on a per-class basis. To achieve this objective, we put forward a novel health criterion indicator, which is developed by integrating various elements, such as Data Preprocessing and Normalization techniques. Additionally, we employ sophisticated techniques like feature selection as well as the utilization of a random forest (RF) classifier. These components are thoughtfully integrated into our approach to create a holistic framework for improving the overall performance and efficiency of the classification process.

The key emphasis of our research lies in the careful selection of optimal features and the incorporation of stability and accuracy metrics for individual classes. This approach is designed to rectify and bridge the gaps in existing methodologies, ultimately leading to a more robust and effective classification process.

The remainder of this paper is organized as follows. The next section provides a detailed description of the structural aspects of the 3 W dataset and the preparation steps undertaken. Section Health criteria indicator construction provides the necessary context for a comprehensive understanding of the proposed methodology. In Section Experimental results and comparative study, we present the experimental results and conduct a comparative study to validate the effectiveness of our approach. Finally, the conclusions of this work are presented in the last section.

## Dataset overview for oil and gas wells

Extracting O&G from under-ground reserves requires a variety of chemical and mechanical operations that can have an impact on the drilling and operation of wells. These operations frequently generate problems that can lead to reduced production or equipment failure. To address these issues, preventive maintenance techniques and regular analysis of production rates, fluids, and mechanical conditions are essential. By maintaining such practices, costly reconditioning and total well losses can be avoided.

## Fault scenarios in oil and gas wells

Petrobras has been developing a comprehensive database since 2010 to document O&G losses within its operational unit in Rio de Janeiro (OU-Rio). This database, known as the loss-integrated management platform, integrates several complementary databases that provide details on production losses through various features. These features include initial and final loss dates, platforms, affected equipment, equipment operators, related sensor tags, original and secondary causes, estimated losses, required actions, subsequent activities, and more.[23]

In a project called "Monitoring of Specialized Alarms", Petrobras launched in mid-2017 to develop an automated system for detecting and classifying eight types of undesirable events in offshore natural flow wells. This project used the 3 W database, described in Section Wrapper feature fusion and selection, and focused on the following event types[6]:

***Scenario* 1**—Abrupt increase of basic sediment and water: This class pertains to the sudden increase in the ratio between water and sediment flow rate and the liquid flow rate, indicating an undesirable amount of water in the produced oil. Early identification of this event is essential to prevent adverse flow assurance conditions, reduced oil production, incrustation, and related issues.

***Scenario* 2**—Spurious closure of the downhole safety valve (DHSV): The DHSV, installed in the production tubing, ensures closure in emergency scenarios or physical disconnection. Detecting spurious closures promptly enables corrective operational procedures to prevent production losses.

***Scenario* 3**—Severe slugging: This class refers to stratified gas-liquid flow scenarios with low liquid and gas flow rates and a declined production line, leading to the accumulation of liquid at the bottom of the riser. The intermittent blockage and subsequent surges of liquid and gas, known as severe slugging, can stress or damage well equipment. Early detection of this event allows preemptive actions to mitigate its effects.

***Scenario* 4**—Flow instability: Flow instability involves sporadic surges of liquid and gas, similar to severe slugging, but less intense and without the complete cycle of liquid blockage followed by a gas surge. If not addressed, flow instability can progress to severe slugging.

***Scenario* 5**—Rapid productivity loss: The productivity of naturally flowing wells depends on various conditions. When the system's energy falls below the minimum required to overcome energy loss, oil flow slows, or stops. Detecting this undesired condition early on minimizes production losses.

***Scenario* 6**—Quick restriction in the production choke (PCK): The PCK controls well operations from the surface. Unintentional quick restrictions, resulting from manual valve operation, directly impact oil production.

***Scenario* 7**—Scaling in PCK: This class relates to the formation of inorganic deposits over time, leading to significant reductions in oil production. Identifying scaling events in their early stages allows for timely intervention, such as injecting scale inhibitors, to prevent further production losses.

***Scenario* 8**—Hydrate in production line: The presence of hydrates in wells and production/injection lines poses a significant challenge in the O&G industry, including Petrobras. Hydrates are crystalline compounds formed by the reaction of natural gas with water, resembling ice. Early detection of hydrate formation helps avoid prolonged production losses, as the process of unblocking production lines is costly and time-consuming.

When operating a well, it is possible for these faults to interact with each other. For example, one fault may trigger another of a different class. For example, flow instability (**scenarios** 4) often precedes severe slugging (**scenarios** 3), as both faults occur in scenarios of low liquid and gas flows and corrugated flow lines. Early detection of these events enables preventive action to be taken to reverse the situation before it becomes critical.

## Structure of database

In this paper, the proposed machine learning CBM system that has been realized and introduced by[6] is used to validate the fault detection and classification performance of the 3 W database. This dataset comprises approximately 2000 operational events, encompassing a range of different well states; this work focuses on detecting 08 realistic faults, which including normal operation (***scenario*** 0) as well as eight distinct fault classes (scenarios 1 to 8). Each event in the dataset consists of time-series data composed of $n = 8$ tags obtained from eight different sensors. The selection of these sensors was based on their availability and relevance to the specific faults being investigated, as outlined in Table 1.

The performance of the proposed controller has been validated in different scenarios by including: real, simulated, and sketched.

- Real validate: are characterized by sensor data collected through the PI System during actual well operations.
- Simulated validate: on the other hand, were generated using the OLGA system,[24] a widely adopted tool in the industry for dynamic multiphase flow simulation.
- Sketched validate: were created using a specially designed tool by the creators of the 3 W database, which leverages expert knowledge to outline the profiles of specific undesirable events.

**Table 1.** Complete tag inventory in the 3 W database: names, descriptions, and units of measurement.

| Sensor name | Sensor description | Measurement unit |
|---|---|---|
| **P-PDG** | Pressure measured at the permanent downhole gauge | Pascal (Pa) |
| **P-TPT** | Pressure measured at the temperature/pressure transducer | Pascal (Pa) |
| **T-TPT** | Temperature measured at the temperature/pressure transducer | Degrees Celsius (°C) |
| **P-MON-CKP** | Upstream pressure of the production choke (CKP) | Pascal (Pa) |
| **T-JUS-CKP** | Downstream temperature of the production choke (CKP) | Degrees Celsius (°C) |
| **P-JUS-CKGL** | Downstream pressure of the gas lift choke (CKGL) | Pascal (Pa) |
| **T-JUS-CKGL** | Downstream temperature of the gas lift choke (CKGL) | Degrees Celsius (°C) |
| **QGL** | Gas lift flow rate | Cubic meters per second (m^3/s) |

To provide a comprehensive understanding of the dataset, Table 2 presents a quantitative breakdown of the 3 W database per events type. It includes the number of occurrences for each ***scenario***, differentiating between real, simulated, and sketched events.

The proposed system aims to improve fault detection and classification performance, ensuring the efficient operation of O&G wells while minimizing potential downtime and related costs.

## Health criteria indicator construction

This section describes the general overview of the different techniques and tools that are used to develop the proposed O&G well's fault detection and classification scheme in term of early detection and highest accuracy with best classification results.

### *Data preprocessing and normalization*

In this sub section, the data preprocessing is an essential step in developing an effective CBM system. Normalization is crucial for standardizing the data and ensuring compatibility across different sensors and measurements, enabling accurate analysis and anomaly detection.

To achieve data normalization, a mathematical equation is utilized to scale the data values between 0 and 1. Let X represents the original data and $X_{norm}$ denotes the normalized data. The normalization equation used in this research is defined as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Here, $X_{min}$ represents the minimum value observed in the original data, and $X_{max}$ denotes the maximum value. Subtracting the minimum value from each data point and dividing it by the range of the data normalizes the values to the desired range.

### *Wrapper feature fusion and selection*

In order to enhance the accuracy and stability of our system, we have developed a novel criterion based on the Wrapper method for feature selection.[25] The objective of this criterion is to select optimal features that improve classification accuracy (referred to as cost1) and enhance stability within each class by reducing the STD (referred to as cost2). Additionally, it aims to reduce the number of input features (referred to as cost3), resulting in improved processing time and faster classification.

Our criterion follows a systematic approach depicted in Figure 1. Initially, a predetermined number of features are selected and utilized by a classifier. The classifier is then simulated N times to evaluate the stability of the classification results. During each simulation, the accuracy and stability of each class are calculated, and this process is repeated N times to obtain average accuracy and STD values.

To quantify the criterion, we employ three cost factors. Cost1 is computed as the sum of weighted accuracy values for each class, represented by $\alpha_i$ and $\mu_i$. Cost2 is calculated as the sum of weighted STDs, represented by $\beta_i$ and $\sigma_i$. Cost3 is determined by dividing the number of selected features by the total number of features.

$$\mathbf{Cost}_1 = \sum_{\mathbf{i}=1}^{\mathbf{C}} \boldsymbol{\alpha}_\mathbf{i}\boldsymbol{\mu}_\mathbf{i} \tag{2}$$

$$\mathbf{Cost}_2 = \sum_{\mathbf{i}=1}^{\mathbf{C}} \boldsymbol{\beta}_\mathbf{i}\boldsymbol{\sigma}_\mathbf{i} \tag{3}$$

$$\mathbf{Cost}_3 = \frac{\textbf{number of Selected feat}}{\textbf{total number of feat}} \tag{4}$$

While

- $$\mu_i = \left(\frac{1}{N}\sum\nolimits_{j=1}^{N} 1 - Acc\ ij\right) \tag{5}$$

- $$\sigma_i = \sqrt{\frac{\sum_{j=1}^{N} (\mathrm{Acc\ ij} - \mu_i)^2}{\mathrm{N}}} \tag{6}$$

Here, N represents the number of tests conducted, and C denotes the number of classes. The accuracy values for

**Table 2.** Quantitative analysis of event types in the 3 W database.

| Class | Description | Real | Simulated | Sketched | Total |
|---|---|---|---|---|---|
| **0** | Normal | 597 | 0 | 0 | 597 |
| **1** | Abrupt BSW increase | 5 | 114 | 10 | 129 |
| **2** | Spurious DHSV closure | 22 | 16 | 0 | 38 |
| **3** | Severe slugging | 32 | 74 | 0 | 106 |
| **4** | Flow instability | 344 | 0 | 0 | 344 |
| **5** | Rapid productivity loss | 12 | 439 | 0 | 451 |
| **6** | Quick PCK restriction | 6 | 215 | 0 | 221 |
| **7** | Scaling in PCK | 4 | 0 | 10 | 14 |
| **8** | Hydrate in Prod. Line | 3 | 81 | 0 | 84 |
| **Total** | | 1025 | 939 | 20 | 1984 |

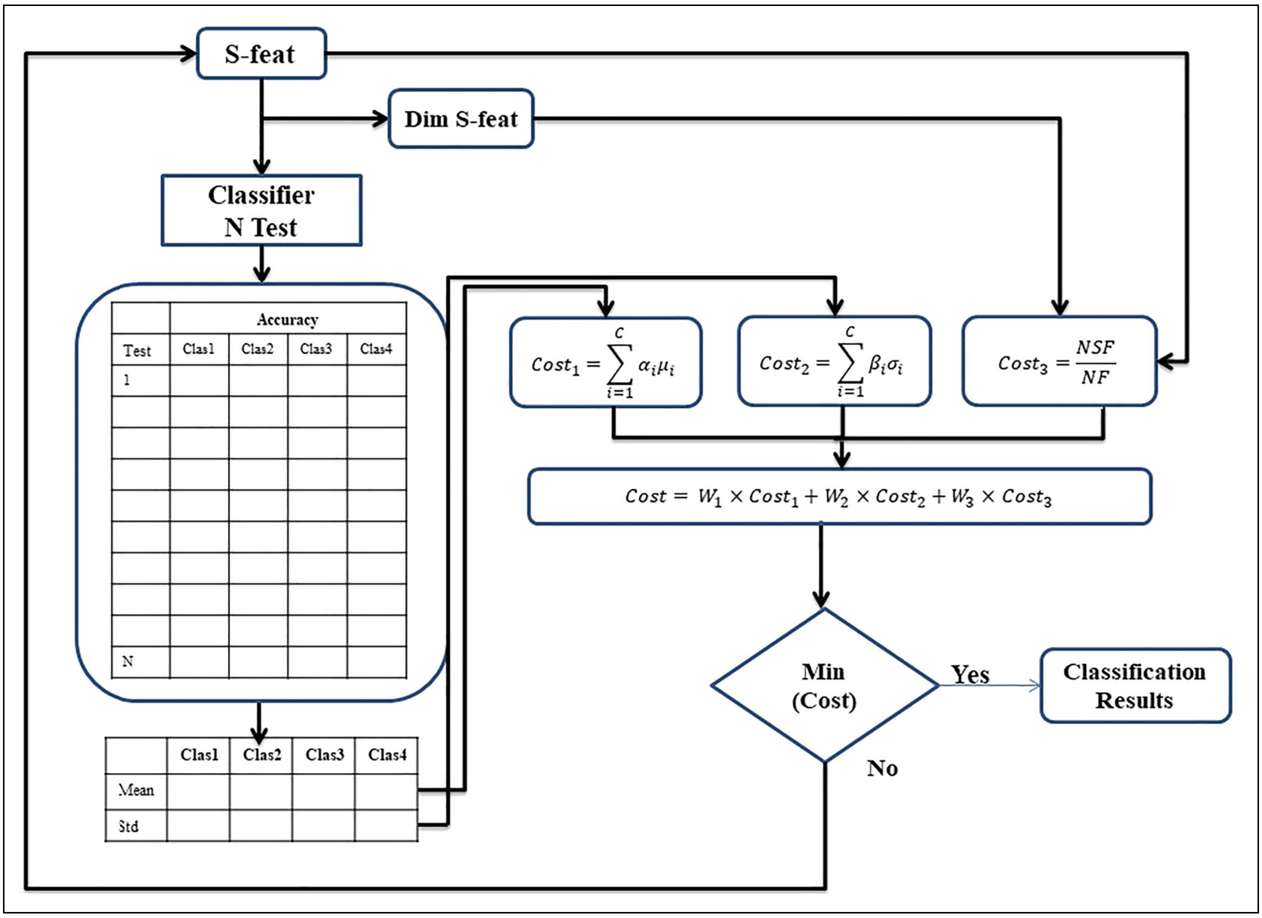


**Figure 1.** Flowchart illustrating the proposed feature selection criteria.[25]

each class in each test are denoted as Acc ij while $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ represent the weights assigned to Cost 1 and Cost 2, respectively. The average accuracy $\mu_i$ and STD$\sigma_i$ are calculated based on the simulation results.

In our cost function, the primary objective is to minimize the error to achieve the highest diagnostic accuracy. The first cost, Cost1, quantifies the error and is calculated as "1 - accuracy." By minimizing this error, we can improve the accuracy of the classification model and obtain better results.

The final cost is determined by combining the results from Cost 1, Cost 2, and Cost 3, utilizing weight values ($\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{W}_3$) assigned to each cost:

$$\mathbf{Cost} = \mathbf{W}_1 \times \mathbf{Cost}_1 + \mathbf{W}_2 \times \mathbf{Cost}_2 + \mathbf{W}_3 \times \mathbf{Cost}_3 \quad (7)$$

The weight assignment is crucial in prioritizing different aspects of the feature selection process. Through experimentation and analysis, we have assigned weights as follows: 0.53 for stability, 0.45 for accuracy, and 0.02 for the number of selected features.

The weight of 0.53 emphasizes the importance of selecting features that exhibit sufficient variance to ensure stability within the classification model. The weight of 0.45 highlights our focus on minimizing errors and achieving high accuracy in the classification process. Lastly, the weight of 0.02 helps control the complexity of the feature set, striking a balance between accuracy, stability, and the number of selected features.

By employing this novel criterion, we aim to improve the overall performance of our system by selecting optimal features that enhance classification accuracy and stability while minimizing the number of input features. This approach contributes to faster processing and improved efficiency in the classification process.

### *Feature classification based on random forest*

After feature fusion and selection, and in order to develop a model capable of predicting defect class and type, we used the RF classifier. RF is a supervised learning and an ensemble technique based on a decision tree N built through bootstrap aggregation, where each tree employs a random sample of the data and each node of the tree is subdivided based on the optimal variable in the subset of input features. This is determined by the Gini index as a measure of attribute selection after calculating the impurity of attributes with respect to classes by equation (8).[26]

$$\textbf{Gini Index} = 1 - \sum\nolimits_{i=1}^{n} \boldsymbol{P}_i^2 \tag{8}$$

where pi denotes an element's probability to be classified for a distinct class.

**Algorithm:** Random Forest.

**Require:** Training Data, Testing Data, Training Labels, Testing Labels, Ntree
**Ensure:** Accuracy
Forest Predictions ← Empty List
**// Create Ntree Number Of Trees**
For I ← 1 To Ntree Do: Sample_I ← Bootstrap Sample (Training Data)
**// Build An Unpruned Classification Tree**
Tree_I ← Build Tree (Sample_I, Training Labels)
**// Predict Using Tree_I On Testing Data**
Predictions_I ← Predict (Tree_I, Testing Data)
**// Add Predictions Of Tree_I To Forest Predictions**
Forest Predictions. Append(Predictions_I)
**// Aggregate Predictions Of All Trees**
Aggregated_Predictions ← Majority Vote (Forest Predictions)
**// Calculate Accuracy**
A ← Size (Aggregated_Predictions)
Correct ← 0
For I ← 1 To A Do:
If Aggregated_Predictions(I) == Testing Labels(I) Then:
Correct ← Correct + 1
End If
End For
Accuracy ← Correct / A
**Return Accuracy**

## Experimental results and comparative study

The process of our proposed approach for early detection and classification of Oil and Gas wells is summarized in the flowchart shown in Figure 2 and it is divided mainly into three steps

### *False alarm detection using binary classification*

The advantage of data normalization is important in the CBM system, as it establishes a consistent scale for all measurements, whatever their original units or ranges. This standardization enables meaningful comparisons and accurate analysis across multiple sensors, contributing to reliable anomaly detection and classification. What's more, standardization reduces the impact of outliers and extreme values by compressing data into a predefined range, thus enhancing system robustness. This enables early detection of deviations and abnormal behavior in well operations, leading to proactive maintenance interventions and optimized production processes.

In this subsection, in our first methodology, we will focus on early detection tasks, specifically examining the false-alarm detection rate and the effectiveness of the selected features. Our objective is to evaluate the performance of our feature fusion and selection approach using the new criteria.

Hence, to ensure the reliability of our analysis, we combined the scenarios 7 from the dataset due to its limited representation of only 14 samples. This decision was made to create a more balanced and robust dataset for binary and multi-class classification. As a result, our revised dataset consisted of 1970 samples, enabling us to mitigate potential biases and derive accurate insights from our analysis.

For the binary classification tasks, we assigned a label of 1 for healthy instances and a label of 2 for faulty instances. This allowed us to assess the false-alarm rate, which represents the rate at which healthy instances are incorrectly classified as faulty. Minimizing the false-alarm detection rate is crucial to ensure the reliability and accuracy of the classification model.

To determine the optimal feature subsets, we employed the RF classifier with several optimization algorithms, the results of classification are shown in Table 3.

Examining the results in the preceding table, which encompasses the assessment of various optimization algorithms alongside the RF classifier, it becomes clear that the Crow Search Algorithm (CSA) excelled in detecting

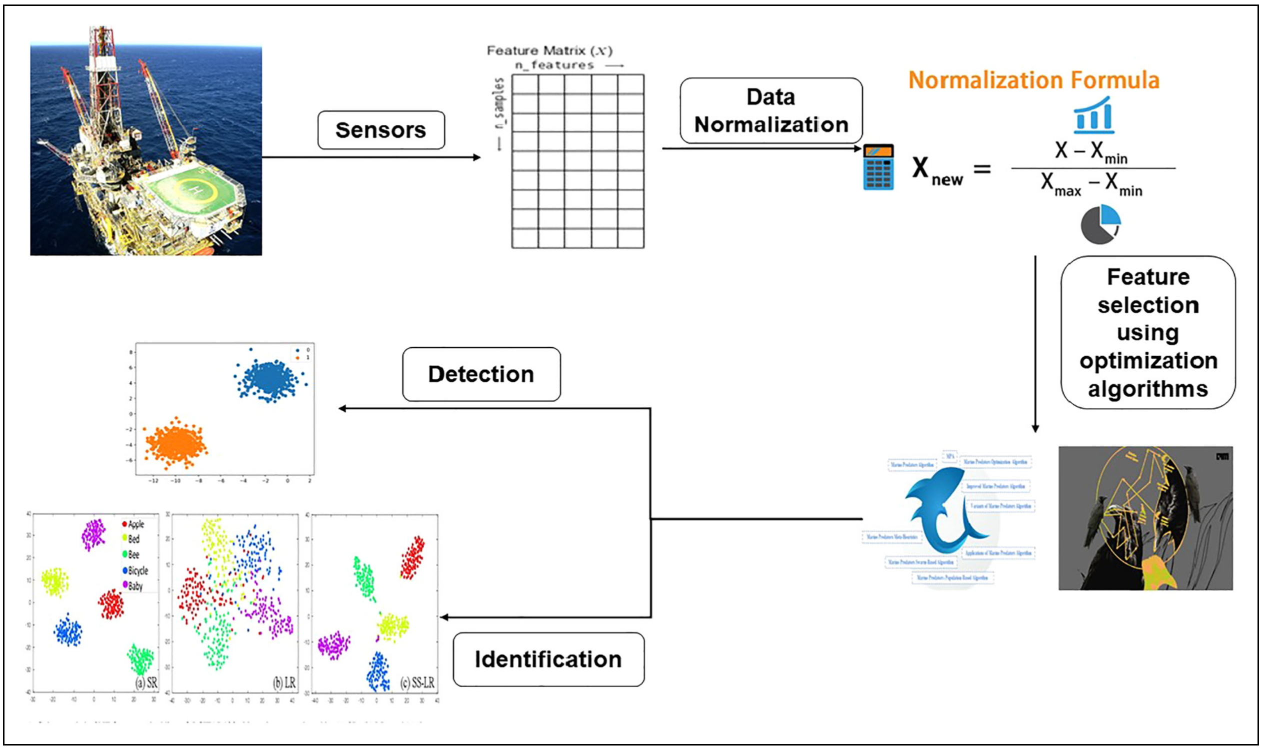


**Figure 2.** Flowchart illustrating the proposed methodology.

**Table 3.** Binary classification results using several optimization algorithms.

| GA | | | ACO | | | CSA | | |
|---|---|---|---|---|---|---|---|---|
| | Scenario 1 | Scenario 2 | | Scenario 1 | Scenario 2 | | Scenario 1 | Scenario 2 |
| Mean | 99.56 | 99.88 | Mean | 99.47 | 99.57 | Mean | 99.63 | 99.85 |
| Max | 100.00 | 100.00 | Max | 100.00 | 100.00 | Max | 100.00 | 100.00 |
| Min | 98.29 | 99.75 | Min | 98.80 | 99.07 | Min | 98.94 | 99.51 |
| STD | 0.58 | 0.13 | STD | 0.44 | 0.29 | STD | 0.36 | 0.17 |
| GNDO | | | MPA | | | TSA | | |
| | Scenario 1 | Scenario 2 | | Scenario 1 | Scenario 2 | | Scenario 1 | Scenario 2 |
| Mean | 99.83 | 99.88 | Mean | 99.34 | 99.73 | Mean | 99.73 | 99.81 |
| Max | 100.00 | 100.00 | Max | 100.00 | 100.00 | Max | 100.00 | 100.00 |
| Min | 98.32 | 99.75 | Min | 98.44 | 99.03 | Min | 98.36 | 99.27 |
| STD | 0.53 | 0.13 | STD | 0.61 | 0.31 | STD | 0.59 | 0.22 |

false alarms. CSA exhibited the lowest STD values across both scenarios and the highest mean accuracy values, underscoring substantial improvements compared to previous findings.[5]

More specifically, in the context of binary classification tasks, CSA achieved a mean accuracy of 99.63% for scenario 1 and an impressive 99.85% for scenario 2. Additionally, CSA exhibited the highest maximum accuracy values for both scenarios, achieving an absolute classification accuracy of 100%. This showcases the proficiency of the CSA-based feature selection approach in accurately classifying instances without triggering false alarms, be it in the healthy scenario 1 or the faulty scenario 2. Furthermore, CSA's minimum accuracy performance underscores the model's resilience, maintaining a high level of accuracy even in the most challenging scenarios, as depicted in Table 3.

Regarding stability, the CSA algorithm showcased exceptional performance, evidenced by a remarkably low STD value. In scenario 1, the STD was a mere 0.36, signifying minimal variability in accuracy results across multiple tests. Likewise, for scenario 2, the STD stood at 0.17, providing further evidence of the algorithm's stability and the consistency of classification outcomes.

These findings unequivocally demonstrate the efficacy of the CSA optimization algorithm when paired with the

RF classifier. It excels in detecting false alarms while concurrently achieving remarkable levels of accuracy and stability in the classification process.

## *Multi scenarios detection using multi-class classification*

To substantiate the effectiveness of our proposed criteria and to enhance the assessment of accuracy and stability across multiple scenarios, we employed these criteria in conjunction with a RF classifier and various optimization algorithms for early detection and classification.

To underscore the robustness of our chosen algorithm, we conducted a comparative study between MPA and five potent nature-inspired optimization algorithms. This study was conducted on a subset of features derived from the extraction of 2 modes across 8 scenarios. Ten (10) simulations were executed for each optimization algorithm, and the outcomes are summarized in Table 4. Our study surpasses the findings reported in a previous research paper, which primarily emphasized overall accuracy without delving into the performance across individual fault scenarios. Moreover, the prior study did not delve into the examination of classification stability, a critical aspect in fault classification.

Comparing the obtained results to the previous paper as shown in Figure 3, we achieved remarkable improvements in both accuracy and classification stability. The RF + MPA combination yielded a mean accuracy of 99.77% for the best-performing fault Scenario and exhibited high accuracy values ranging from 94.34% to 98.94%. Moreover, our evaluation of classification stability revealed low STD values (ranging from 0.00 to 3.07) for each fault Scenario. This demonstrates the consistency and reliability of our classification results, which is a significant advantage over the previous study that overlooked this aspect.

As result, the proposed study not only surpasses the overall accuracy reported in the previous paper but also addresses the limitations by evaluating classification performance for individual fault Scenarios and assessing

**Table 4.** Comparison between MPA, TSA, CSA, ACO, GNDO, and GA optimization algorithms with RN classifier.

| Random forest in tandem with tree-seed algorithm TSA | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 | Scenario 6 | Scenario 7 | Scenario 8 |
| Mean | 99.45 | 96.29 | 99.23 | 98.04 | 99.9 | 99.12 | 97.14 | 97.83 |
| STD | 0.59 | 2.86 | 2.43 | 2.24 | 0.32 | 0.65 | 1.06 | 2.53 |
| Random forest in tandem with Marine Predators Algorithm MPA | | | | | | | | |
| | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 | Scenario 6 | Scenario 7 | Scenario 8 |
| Mean | 99.77 | 97.11 | 100 | 94.34 | 99.9 | 98.44 | 98.94 | 97.84 |
| STD | 0.4 | 2.59 | 0 | 2.76 | 0.33 | 1.28 | 1.21 | 3.07 |
| Random forest in tandem with Generalized normal distribution optimization GNDO | | | | | | | | |
| | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | scenario5 | Scenario 6 | Scenario 7 | Scenario 8 |
| Mean | 99.78 | 98.09 | 92.59 | 97.59 | 99.58 | 99.64 | 99.15 | 97.59 |
| STD | 0.47 | 1.85 | 6.04 | 3.36 | 0.74 | 0.38 | 1.34 | 3.54 |
| Random forest in tandem with Ant colony optimization ACO | | | | | | | | |
| | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 | Scenario 6 | Scenario 7 | Scenario 8 |
| Mean | 99.79 | 97.49 | 93.91 | 94.08 | 99.69 | 99.04 | 97.86 | 98.44 |
| STD | 0.43 | 2.73 | 5.7 | 5.06 | 0.71 | 0.91 | 1.21 | 2.02 |
| Random forest in tandem with Genetic algorithm GA | | | | | | | | |
| | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 | Scenario 6 | Scenario 7 | Scenario 8 |
| Mean | 99.5 | 96.57 | 94.68 | 94.62 | 99.5 | 99.19 | 99.38 | 96.6 |
| STD | 0.6 | 1.54 | 6.4 | 4.09 | 0.86 | 1.14 | 1.46 | 3.53 |
| Random forest in tandem with Crow search Algorithm CSA | | | | | | | | |
| | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 | Scenario 6 | Scenario 7 | Scenario 8 |
| Mean | 99.95 | 97.28 | 89.8 | 96.27 | 99.63 | 99.43 | 98.89 | 97.62 |
| STD | 0.17 | 3.62 | 10.68 | 4.46 | 0.48 | 0.65 | 1.71 | 2.11 |

classification stability. These advancements contribute to a more accurate and reliable fault classification system.

To perform an analysis of the defect recognition ability of the proposed method for ***Multi scenarios detection***, a set of faults predictions on 08 scenarios faults of the test set is implemented. The confusion matrix of the prediction results is shown in Figure 4.

It can be seen from Figure 4, of scenario 1, and scenario 6 our predicted model reaches an impressive accuracy, respectively, 99.44%, and 99.17% with a neglected misclassification rate of 0.56%. Similarly, for the scenario 2, and scenario 4, with minimal classification error rates, the accuracy of predicted model reaches a substantial 97.83% and 96.55% respectively. However, the most notable achievement is observed in scenario 3, where the predicted model achieves a very high accuracy of 100%.

The obtained results reflect the model's ability to effectively handle the complexities associated with these fault types, further improving the overall performance of our system. In addition, the proposed predicted model successfully achieves with a higher accuracy of 100% for scenarios 5, 7, and 8, demonstrating its ability to accurately detect these fault types without any mis-classifications.

Finally, we concluded that the comparison with the previous research paper is particularly compelling, as they only considered overall accuracy without assessing classification stability for each scenario. By integrating the new criteria and evaluating stability meticulously, the predicted model provides valuable

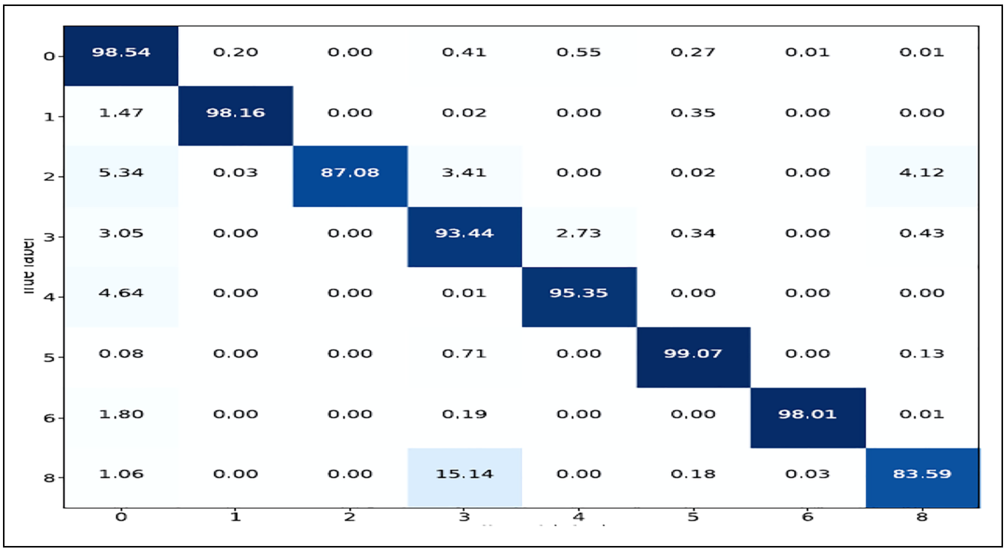


**Figure 3.** Confusion matrix of the previous research paper.[5]

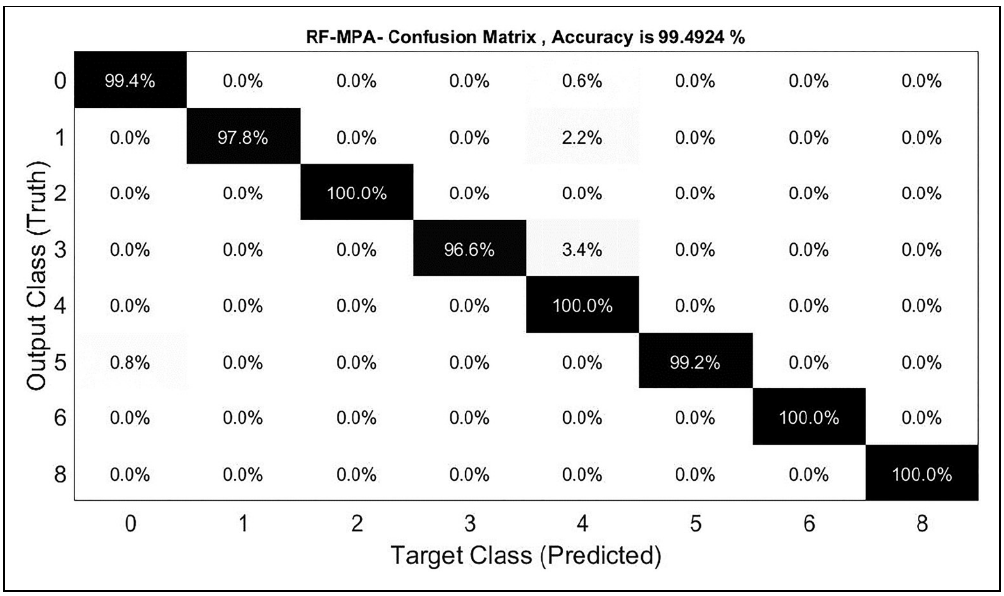


**Figure 4.** Classification results using our new health criteria.

information on classification performance for individual fault scenarios.

## Conclusion

In conclusion, this research paper marks a significant milestone in the realm of fault classification within the O&G industry. Our primary objective was to improve upon existing methodologies by introducing innovative criteria and a comprehensive evaluation of classification stability, thereby enhancing both accuracy and reliability. Our findings demonstrate a notable advancement in fault detection and classification techniques, far surpassing the outcomes of a prior research publication. Through meticulous experimentation and rigorous analysis, our proposed MPA + RF model has proven its exceptional prowess in accurately identifying fault types. The mean accuracy metrics for each class, complemented by our thorough evaluation of classification stability, provide a robust foundation for practical applications.

In direct comparison to a previous study, our approach not only outperforms their reported results but also introduces a new dimension of reliability in fault classification. By incorporating novel criteria and conducting an exhaustive assessment of stability, we have achieved superior accuracies and fortified the credibility of fault classification results. The examination of the confusion matrix further reinforces the outstanding accuracy of our model, with an exceptional 100% accuracy rate for Class 3, showcasing its precision in fault identification. Our model consistently excels across various fault classes, surpassing the accuracy levels reported in the prior research paper, thus reaffirming its superiority. The quantitative results, supported by rigorous scrutiny of mean accuracies and STDs, attest to the precision and consistency of our model's predictions. These quantitative outcomes underscore the practical applicability of our approach, providing a solid foundation for its deployment in real-world scenarios.

Looking ahead, the potential for future research is vast. Optimizing our model for scalability in large-scale applications, integrating real-time data streams, and exploring advanced anomaly detection techniques are avenues ripe for exploration. In summary, our study not only contributes to the advancement of fault detection in the O&G sector but also paves the way for future research endeavors that will undoubtedly push the boundaries of accuracy and reliability in this critical domain.

### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Mohammed Amine Sahraoui https://orcid.org/0009-0005-9435-1374
Chemseddine Rahmoune https://orcid.org/0000-0003-0388-7455
Fawzi Gougam https://orcid.org/0000-0001-6964-1867

### References

1. Liu J, Hou L, Zhang R, et al. Explainable fault diagnosis of oil-gas treatment station based on transfer learning. *Energy* 2023; 262: 125258.
2. Jovicic E, Primorac D, Cupic M, et al. Publicly Available Datasets for Predictive Maintenance in the Energy Sector: A Review. 2023. doi:10.1109/ACCESS.2023.3295113.
3. Wang Y, Estefen SF, Lourenço MI, et al. Optimal design and scheduling for offshore oil-field development. *Comput Chem Eng* 2019; 123: 300–316.
4. Wei C, Tabjula JL, Sharma J, et al. A novel data assimilation-based real-time state estimation method for gas influx profiling during riser gas events. *J Energy Resourc Technol* 2023: 1–39.
5. Marins MA, Barros BD, Santos IH, et al. Fault detection and classification in oil wells and production/service lines using random forest. *J Pet Sci Eng* 2021; 197: 107879.
6. Vargas REV, Munaro CJ, Ciarelli PM, et al. A realistic and public dataset with rare undesirable real events in oil wells. *J Pet Sci Eng* 2019; 181: 106223.
7. Magnusson LV, Olsson JR and Tran CTTJIIS. Recurrent neural networks for oil well event prediction. *IEEE Intell Syst* 2023; 38: 73–80.
8. Van Oort E, Griffith J and Schneider B. How to accelerate drilling learning curves. In: *SPE/IADC Drilling Conference and Exhibition*. OnePetro, 2011.
9. Kemajou V and Samuel R. Real-time well monitoring and engineering analysis of drilling activities: intelligent rig state detection and prediction with uncertainty. In: *International conference on offshore mechanics and Arctic engineering*. American Society of Mechanical Engineers, 2020, pp.V011T011A001.
10. Pournazari P, Ashok P, van Oort E, et al. Enhanced kick detection with low-cost rig sensors through automated pattern recognition and real-time sensor calibration. In: *SPE Middle East Intelligent Oil and Gas Conference and Exhibition*. OnePetro, 2015.
11. Osman A and Ovinis MJM. A review of in-situ optical flow measurement techniques in the deepwater horizon oil spill. *Measurement* 2020; 153: 107396.
12. Beyer J, Trannum HC, Bakke T, et al. Environmental effects of the deepwater horizon oil spill: a review. *Mar Pollut Bull* 2016; 110: 28–51.
13. Upton HF. The Deepwater Horizon oil spill and the Gulf of Mexico fishing industry. In: 2011, Congressional Research Service, Library of Congress.
14. Johnston B, Foley A, Doran J, et al. Levelised cost of energy, A challenge for offshore wind. *Renewable Energy* 2020; 160: 876–885.
15. Smith K. *Environmental hazards: assessing risk and reducing disaster*. Routledge, 2013.
16. Chalapathy R and Chawla S. Deep learning for anomaly detection: a survey. *Comput Sci* 2019.

17. Abbasi T, Lim KH and Yam KS. Predictive maintenance of oil and gas equipment using recurrent neural network. In: *Iop conference series: materials science and engineering*. IOP Publishing, 2019, pp.012067.
18. Gatta F, Giampaolo F, Chiaro D, et al. Predictive maintenance for offshore oil wells by means of deep learning features extraction. *Expert Syst* 2022: e13128.
19. Machado APF, Vargas REV, Ciarelli PM, et al. Improving performance of one-class classifiers applied to anomaly detection in oil wells. *J Pet Sci Eng* 2022; 218: 110983.
20. Turan EM and Jäschke J. Classification of undesirable events in oil well operation. In: *2021 23rd international conference on process control (PC)*. IEEE, 2021, pp.157–162.
21. Aslam N, Khan IU, Alansari A, et al. Anomaly detection using explainable random forest for the prediction of undesirable events in oil wells. *Appl Comput Intell Soft Comput* 2022; 2022.
22. Castro AODS, Santos MDJR, Leta FR, et al. Unsupervised methods to classify real data from offshore wells. *Am J Oper Res* 2021; 11: 227–241.
23. Santos IH, Lisboa H, Feital T, et al. Hydrate failure detection in production and injection lines using model and data-driven approaches. *Rio Oil Gas Expo Conf* 2018.
24. Schlumberger O. Dynamic Multiphase Flow Simulator. Version, 2014.
25. Sahraoui MA, Rahmoune C, Meddour I, et al. New criteria for wrapper feature selection to enhance bearing fault classification. *Adv Mech Eng* 2023; 15: 168781322311 83862.
26. Imane M, Rahmoune C and Benazzouz D. Rolling bearing fault feature selection based on standard deviation and random forest classifier using vibration signals. *Adv Mech Eng* 2023; 15: 16878132231168503.

## Appendix

### *Abbreviations*

| | |
|---|---|
| ARAD | Automatic rig-activity detection |
| ML | Machine learning |
| O&G | oil and gas |
| CBM | Condition-based Monitoring |
| STD | Standard deviation |
| BSW | Basic sediment and water |
| DHSV | Downhole safety valve |
| PCK | Production choke |
| P-PDG | Pressure measured at the permanent downhole gauge |
| P-TPT | Pressure measured at the temperature/pressure transducer |
| T-TPT | Temperature measured at the temperature/ pressure transducer |
| P-MON-CKP | Upstream pressure of the production choke (CKP) |
| T-JUS-CKP | Downstream temperature of the production choke (CKP) |
| P-JUS-CKGL | Downstream pressure of the gas lift choke (CKGL) |
| T-JUS-CKGL | Downstream temperature of the gas lift choke (CKGL) |
| QGL | Gas lift flow rate |
| RF | Random Forest |
| CSA | Crow Search Algorithm |
| ACO | Ant Colony optimization |
| GA | Genetic Algorithm |
| GNDO | Generalized Normal Distribution Optimization |
| MPA | Marine Predators Algorithm |
| TSA | Tree-Seed Algorithm |