A System to Detect Oilwell Anomalies Using Deep Learning and Decision Diagram Dual Approach

P. E. Aranha[1,2]*, L. G. O. Lopes[3], E. S. Paranhos Sobrinho[3], I. M. N. Oliveira[3], J. P. N. de Araújo[3], B. B. Santos[3], E. T. Lima Junior[3], T. B. da Silva[3], T. M. A. Vieira[3], W. W. M. Lira[3], N. A. Policarpo[2], and M. A. Sampaio[2]

[1]Well Construction Department, Petrobras
[2]Department of Mining and Petroleum Engineering, Universidade de São Paulo
[3]Laboratory of Scientific Computing and Visualization, Federal University of Alagoas

Summary

Detecting unexpected events is a field of interest in oil and gas companies to improve operational safety and reduce costs associated with nonproductive time (NPT) and failure repair. This work presents a system for real-time monitoring of unwanted events using the production sensor data from oil wells. It uses a combination of long short-term memory (LSTM) autoencoder and a rule-based analytic approach to perform the detection of anomalies from sensor data. Initial studies are conducted to determine the behavior and correlations of pressure and temperature values for the most common combinations of well valve states. The proposed methodology uses pressure and temperature sensor data, from which a decision diagram (DD) classifies the well status, and this response is applied to the training of neural networks devoted to anomaly detection. Data sets related to several operations in wells located at different oil fields are used to train and validate the dual approach presented. The combination of the two techniques enables the deep neural network to evolve constantly through the normal data collected by the analytical method. The developed system exhibits high accuracy, with true positive detection rates exceeding 90% in the early stages of anomalies identified in both simulated and actual well production scenarios. It was implemented in more than 20 floating production, storage, and offloading (FPSO) vessels, monitoring more than 250 production/injection subsea wells, and can be applied both in real-time operation and in testing scenarios.

Introduction

For decades, the oil and gas industry has been a key player in the global economy by meeting the world's energy demand. According to Zhong et al. (2020), fossil sources currently satisfy almost 75% of the world's total demand and are expected to remain the primary energy source for several more years. To improve efficiency and safety, the industry is evolving and embracing digital transformation by incorporating advanced technologies such as artificial intelligence (AI), cloud computing, automation, big data, and the Internet of Things.

The oil and gas industry processes are highly complex, requiring careful execution throughout the entire chain, from exploration to distribution. In this context, the adoption of AI methodologies stands out as a game-changing initiative to ensure safety, robustness, and predictability. Gurina et al. (2020) developed an anomaly detection approach for directional drilling operations, while Anifowose et al. (2019) compared traditional and sophisticated machine learning techniques for predicting petroleum reservoir permeability. Zhao et al. (2022) combined machine learning and numerical simulation to model fracture propagation in shale formation. Wang and Chen (2019) utilized LSTM networks to learn the pressure transient behavior in naturally fractured tight reservoirs. In Tang et al. (2021), generative adversarial networks (GANs) are used along with gradient boosting algorithms to predict shale reservoir quality in searching for sweet spots.

The smooth and safe production of oil wells requires continuous attention to ensure that the fluid from the reservoir reaches the surface efficiently. However, due to varying temperature and pressure conditions throughout the well's life cycle, issues such as equipment failure, poorly performed operations, or unforeseen circumstances during the design stage may arise. As a result, monitoring equipment is installed to observe the behavior of the well. According to Hüffner et al. (2019), pressure and temperature sensors are typically located at the bottom of the production column, specifically the permanent downhole gauge (PDG), and at the wet Christmas tree, in the case of the temperature/pressure transducer (TPT). The data transmitted from these sensors can be used to assess the condition of the well.

In many industries, including the oil industry, monitoring various data of interest through sensors is a common practice to detect possible undesired conditions in structures and prevent accidents. Vargas et al. (2019) emphasized that detecting and classifying rare undesirable events is a crucial task in various activities monitored by humans. Venkatasubramanian (2003) added that the mission of responding to unexpected events involves promptly detecting anomalous behaviors, diagnosing their root causes, and taking appropriate control actions to bring operations back to a healthy, safe, and operational state.

To achieve autonomous monitoring, various intelligent techniques are applied, including machine learning methodologies and, more specifically, deep learning. These latter can be particularly useful in identifying possible anomalies in the data collected by sensors. Marchi et al. (2015) and Malhotra et al. (2015) proposed methodologies that use deep learning to detect abnormal behavior in time series data. Cheng et al. (2022) applied deep learning techniques for the prediction and detection of scale buildup in production wells. A systematic review on the use of AI in oil well drilling and production is provided by D'Almeida et al. (2022).

Machine learning techniques have been applied to well production monitoring for the detection of specific events. Alharbi et al. (2022) addressed anomaly detection in production data, comparing the performance of models that they refer to as white box and black box, with respect to their explainability. Among the models tested, decision tree (DT) stands out for its interpretability and performance. Turan and Jaschke (2021) applied different classifiers such as adaptive boosting (AdaBoost), linear and quadratic discriminant analysis, DT, and

*Corresponding author; email: pearanha@petrobras.com.br

Original SPE manuscript received for review 27 April 2023. Revised manuscript received for review 6 October 2023. Paper (SPE 218017) peer approved 23 October 2023.

random forest (RF) to the analysis of well production problems presented in Vargas et al. (2019). The authors stated that DT achieved a balanced accuracy of 88%, performing better than more robust techniques such as RF and AdaBoost. Figueirêdo et al. (2021) applied six unsupervised models to detect anomalous events in a production line, and their best-ranked model achieved an accuracy ranging from 97% to 100%. In addressing the scale deposition problem, Yousefzadeh et al. (2023) compared the performance of 10 different intelligent algorithms, highlighting the performance of K-Nearest Neighbors and ensemble learning models. Al-Hajri et al. (2020) used classification techniques such as DT and gradient boosting classifier to generate prediction results for scale buildup, which are used as inputs for a probabilistic inhibition-design model. Monday and Odutola (2021) used both supervised and unsupervised learning algorithms to address hydrate formation, while Gjelsvik et al. (2023) conducted a systematic review on the use of intelligent techniques to tackle the gas hydrate problem.

Various models based on support vector machines (SVMs), LSTM, and GANs have been presented in the literature for detecting anomalous behavior from pressure and temperature data gathered from subsea well sensors. These models can recognize undesirable events based on the short- and long-term responses used for training. Time-series novelty detection, or anomaly detection, is a challenging subject in data mining that refers to the automatic identification of new or abnormal events incorporated in the time-series regular points. This subject is gaining increasing attention due to its enormous potential for immediate applications. One-class SVMs are a well-established method in the literature, as demonstrated in the works of Vargas et al. (2019), Ma and Perkins (2003), and Shawe-Taylor and Zlicar (2015). The RF algorithm is also used for anomaly detection, as can be seen in the study by Marins et al. (2021), who address different well production problems presented in the work of Vargas et al. (2019).

LSTM networks were proposed to create more robust recurrent neural networks that avoid vanishing gradients. They have a sophisticated formulation that enables selective "remember and forget" configurations, allowing for the determination of the amount of previous data to be considered for predicting the next data. Several studies such as Vávra and Hromada (2019), Marchi et al. (2015), and Malhotra et al. (2015) have presented methodologies for detecting abnormal behavior in time-series data by using only data from systems in regular operation. In Machado et al. (2022), LSTM and one-class SVM were applied successfully to detect two short- and long-term problems in well production lines, namely, spurious valve closure and hydrate formation. The results are compared with tree-based classifiers such as DT and RF models.

GANs use data from standard instances to train two interconnected neural networks—the generator and the discriminator. The generator reproduces typical operating scenarios, and the discriminator can detect anomalies and classify the input data. Recent studies, such as Li et al. (2018), Sabokrou et al. (2018), and Mattia et al. (2019), have shown that using the GAN discriminator to detect abnormal system behavior produces good results.

The three anomaly detection methodologies discussed have unique characteristics that differentiate them. SVMs have pre-established configurations that make them easy to implement but lack the versatility of the other two methods. While they achieve an accuracy of 70–90% in detecting abnormal states according to Marins (2018), changing the tolerance degree requires adjusting the SVM parameters, which is time-consuming. GAN-based networks are the most sophisticated approach but require a complex configuration and a significant amount of data for training. Although they are capable of accurately reproducing complex signals, their requirement for a robust computer infrastructure can make them impractical for real-time or embedded systems.

In contrast, LSTM networks are simple to train and can automatically configure for each well and valve state, being versatile in adjusting their configuration and tolerance for detecting anomalies. It is worth mentioning its ability to handle sequential data, considering nonlinear and long-term dependencies, as it is observed in the pressure and temperature series collected. These data sets can exhibit complex patterns due to the evolving conditions of the reservoir and the well over time. The memory capability of LSTMs allows for the detection of anomalies that occur intermittently or over extended periods. Regarding the specific problem addressed in this paper, there is a limited availability of labeled anomalous data due to the rare occurrence of anomalies, which is the rationale for choosing an LSTM-based autoencoder, which can be trained using only data labeled as normal.

With that said, the current work proposes a comprehensive anomaly detection methodology that combines LSTM autoencoders and a rule-based analytical approach to real-time anomaly detection in subsea well sensor pressure and temperature data. It is worth mentioning that the methodology relies solely on normal data for training and is capable of detecting anomalies, whether they have been previously mapped or not. It was integrated into a monitoring framework, and this system has been able to deliver a prompt response in online well integrity monitoring, improving operational safety and reducing costs associated with NPT and failure repair in production/injection wells.

Methodology

Sensors installed to monitor production in oil wells are located at a few viable positions to monitor the behavior of a well during its operation and regularly transmit data concerning pressure and/or temperature values along the depth. With an adequate treatment of these data, followed by pattern identification techniques, it becomes possible to evaluate if the well operates within normality or not, allowing operators and engineers to be capable of classifying normal and abnormal situations by analyzing the time series.

In the anomaly detection system, there is an entry point for the sensor data, from which checks are performed with the association of a DD and a deep learning network, as illustrated in **Fig. 1.** The system can map valve configurations and set its anomaly search mechanisms to suit the current valve status. The methodology is developed in such a way that the system is scalable to monitor simultaneously multiple wells.

The anomalies detected by the DD can be classified as mapped and unmapped. For the first case, the tool has a portfolio of anomalies that contains the attributes of undesired scenarios in terms of pressure and temperature. The portfolio was designed based on the expected behavior of the well for a given valve configuration. Therefore, the attributes of each scenario are defined by rules and tolerances. As for the second group of anomalies, these are the cases that have not yet been mapped in the system but configure a scenario that does not correspond to the normality of the well in question, being detected due to the behavior patterns that the system attributes to the well for both analytical DD and deep learning approaches.

Analysis of Pressure and Temperature Sensor Data. One of the steps in data processing is grouping and synchronizing the pressure and temperature sensor data provided by the oil company. The time series of each sensor is stored in files, containing information gathered in different time intervals. The data were synchronized and grouped based on combinations of valve states in production. The first analysis considered the two main well valve combinations as output, which differ from each other by the open or closed state of the Master 2 valve (M2). Correlation studies are carried out to identify behaviors of well operation that differ from normal. **Fig. 2** illustrates the correlation

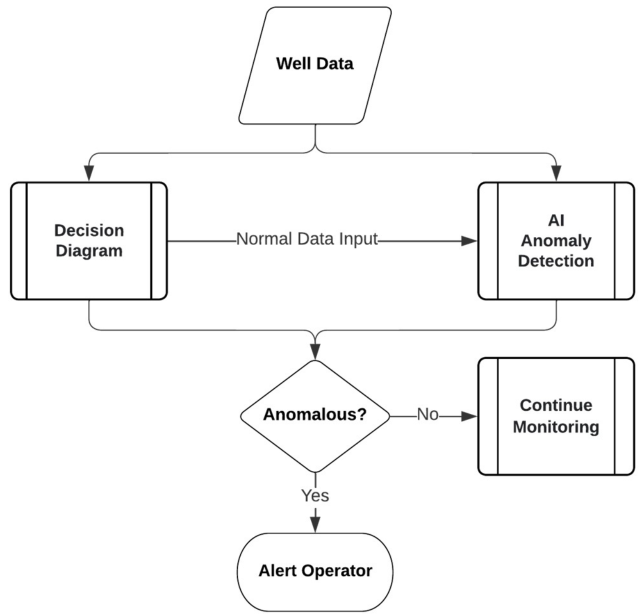

Fig. 1—General methodology of the proposed dual process.

between pressure and temperature values collected from sensors, comparing data from the literature (Vargas et al. 2019) with industry data provided for both the open and closed M2 valves.

The red color indicates a positive correlation (directly related variables), and the blue color indicates a negative correlation (inversely proportional variables). This study indicates that different valve status configurations should be investigated and taken into consideration during the development of methodologies for anomaly detection, as they influence the correlation between variables.

Anomaly Detection Using the DD. The methodology that defines an action to be taken by the DD consists of structuring a set of rules based on layers with nodes. Each node verifies a condition that determines the next node that the information will be passed to, until

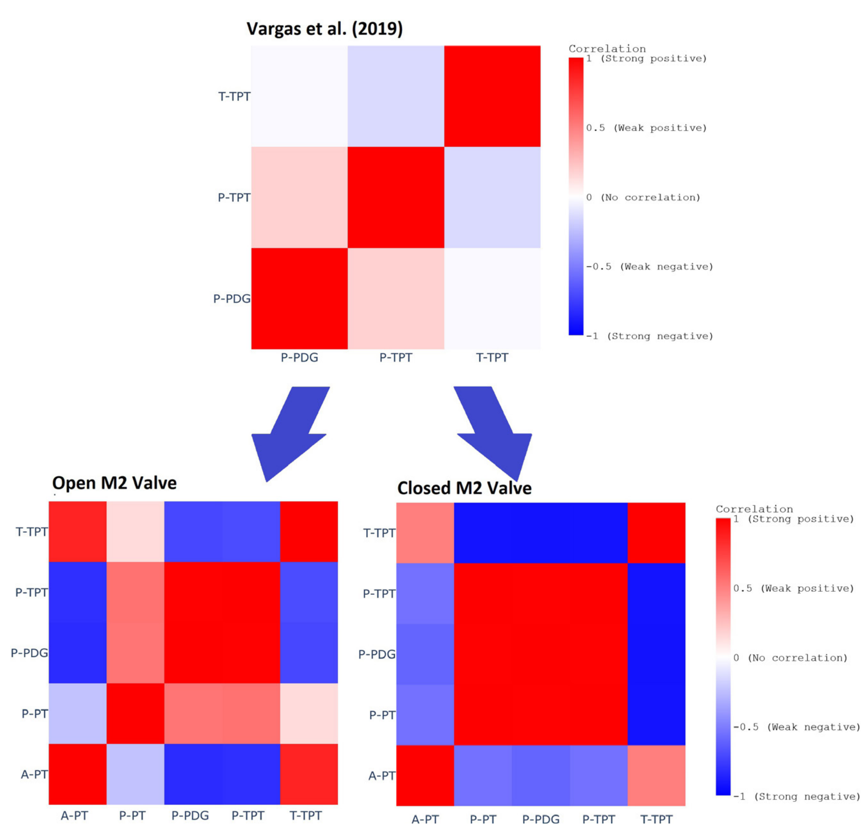

Fig. 2—Correlation matrices for sensor data.

it reaches a leaf node. As decisions advance through the internal nodes, the data become increasingly specific. As a result, the DD can effectively manage the information to extract maximum benefits and indicate the occurrence of anomalies.

The first layer of the DD refers to the data of the monitored well, given the tool's capability to track multiple wells simultaneously. In the next layer, the type of well data is checked, determining if it is the initial data or if a valve change has occurred. In the case of initial well data, only the well data structure is created and stored in the database. However, if there has been a change in the state of the operating valves, the system enters standby mode until pressure and temperature levels are normalized. When the state of the valves remains the same, the system performs well behavior analysis to identify possible anomalies.

Finally, the well's state is reported, which can be categorized as either Normal when the pressure and temperature levels are within the expected range, or Analyzing when there is a change in the status of a valve, avoiding false alerts when the sensors are still stabilizing after a valve closure/opening. In addition to these, there are anomaly statuses, such as Anomalous, for when there is a significant change in pressure and temperature levels (greater than a previously set threshold), and the cases of M1/DHSV Closed, W1 Closed, M2 Closed, and Tubing leakage. For example, in the case of a production well, if the M1/DHSV valves are supposed to be open according to the monitoring system, but the pressure at the PDG increases while the pressure at the TPT decreases, the DD alerts that the M1/DHSV valves might be closed. Another scenario is when the W1 valve is reported as open, but there is a sudden simultaneous increase in pressure readings from the TPT and PDG sensors with the same choke opening. In such a case, the monitoring system triggers an alert, suggesting that the W1 valve might be closed. **Fig. 3** displays the structure of the DD.

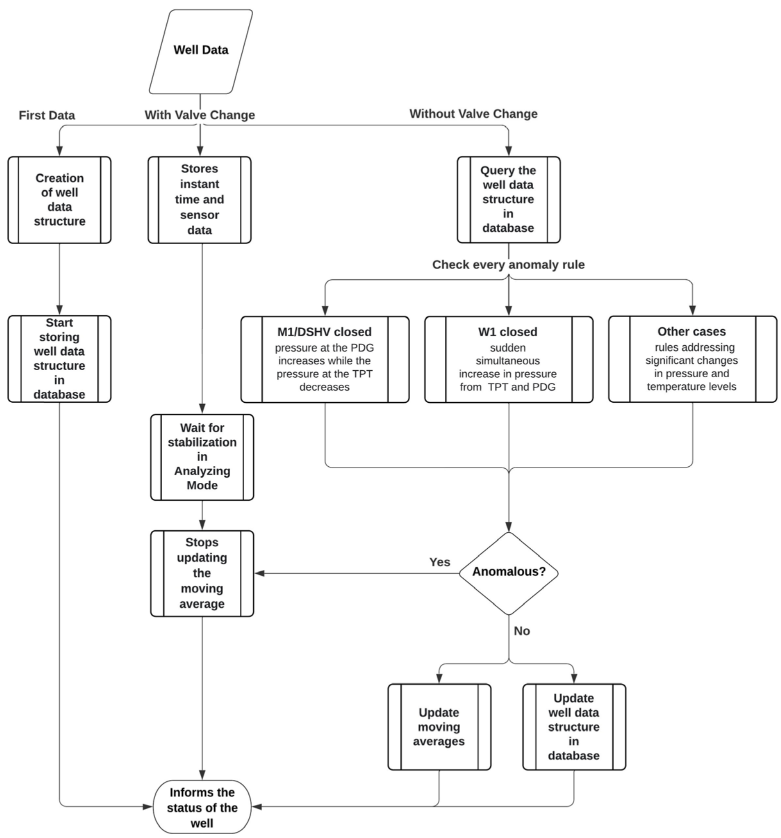

Fig. 3—Structure of the DD.

The data concerning pressure and temperature sensors PDG, TPT, and PT are first analyzed together with the state of the monitored valves—DHSV, M1, M2, W1, W2, XO, PXO and Choke valve—to develop the proposed methodology. The typical valve and sensors scheme for a subsea oil well is described in **Fig. 4.** The data history is collected from wells and production vessels (FPSOs) in different oil basins. Initial studies are conducted to determine the behavior and correlations from the pressure and temperature data for the most common combinations of valve states. An analytical rule-based approach is then developed to detect abnormal variations in the data, serving as a basis for classifying the anomaly based on a known portfolio.

The response of this analytical system is applied to neural network training, which is also dedicated to anomaly detection. The trained models receive real-time data and monitor valve behavior using load conditions that may indicate abnormal events during oil production based on the current valve scheme. Multiple data sets from the literature (Vargas et al. 2019) were used to train and validate the proposed methodology.

After validating the joint application of the analytical and machine learning methodologies, an interface is developed using the Node-Red platform coupled with the Python 3.7 code. This tool can receive data from different wells and perform simultaneous monitoring using advanced methods.

Anomaly Detection Through Machine Learning. An important challenge for anomaly detection in the oil industry is that data of anomalies are scarce, so it is difficult to train an AI model accurately enough to classify anomalies. So, the proposed methodology involves automatically learning what fits as a normal operation state, which happens most of the time, and constantly checking if the new data match the previously trained. Once the system has learned what is considered normal, it can alert the operator when it detects an abnormal state. This means that, for each oil well, the network must collect data and learn on its own without requiring operator interference. In

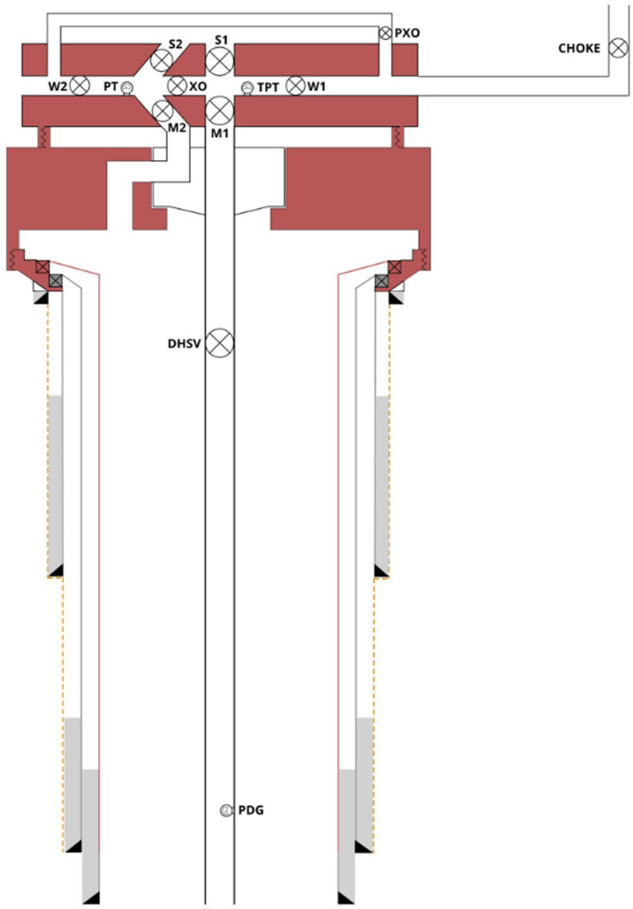

Fig. 4—Oilwell valves and sensors scheme.

this sense, two modules are developed, one for learning and the other for detection. The learning module stores sequential temporal data classified as normal by the DD formulation. The module enters the training mode after acquiring approximately 500 normal data points, which covers roughly 5,000 seconds of real-time production, an amount determined through tests using industry data. It generates three outputs: (1) data normalization parameters, (2) neural network weights, and (3) anomaly tolerance calculated during training. After saving these three outputs, the detection module starts operating, detecting abnormal behaviors in newly acquired data. Networks based on different configurations were tested, and the LSTM-based autoencoder proposed by Larzalere (2019) had the best results.

The LSTM autoencoder input consists of multivariate time-series data capturing temperature and pressure readings from various sensors. Given that the model is an autoencoder, it is trained to reconstruct the input time series after passing through several bottleneck hidden layers, which first encode the original data in a lower-dimension space and then decode the original data. These reconstructed data are compared with the original input, yielding what we term the reconstruction error. This error is then measured against a predefined threshold set during training. Consequently, the LSTM output categorizes the result as either Normal or Anomalous.

Fig. 5 depicts the network developed, in which None refers to the batch size, that is not predefined on the model's constructor, and ReLU refers to the rectified linear activation function. The LSTM features consist of pressure values from the sensors TPT, PT, and PDG, as well as the temperature values from TPT and PDG.

The Adam optimizer (Kingma and Ba 2014), available in the Keras framework (Chollet 2015), is used to calculate the network weights. The Adam algorithm enhances the classical stochastic gradient descent procedure for optimization. Regarding the error metric, mean absolute error and symmetric mean absolute percentage error (SMAPE) were applied and obtained similar results, leading to the choice of SMAPE as the main error metric, which is given by Eq. 1:

$$\mathrm{SMAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\frac{\left|F_t + A_t\right|}{\left(\left|A_t\right| + \left|F_t\right|\right)/2}, \tag{1}$$

where A_t is the real value and F_t is the model predicted value.

In the studies, the number of epochs is 500 with a batch size of 100. Using a similar approach, the training of the deep network uses 500 instances of data labeled as normal in the DD. After training, the SMAPE value for each prediction is calculated, then approaching this error metric as a normal distribution of values.

Assuming that the reconstruction error follows a Gaussian distribution, the tolerance is evaluated by Eq. 2, in terms of the mean of the parameter and its standard deviation, defined as μ and σ, respectively:

$$\mathrm{tol}_{AI} = \mu + k\sigma, \tag{2}$$

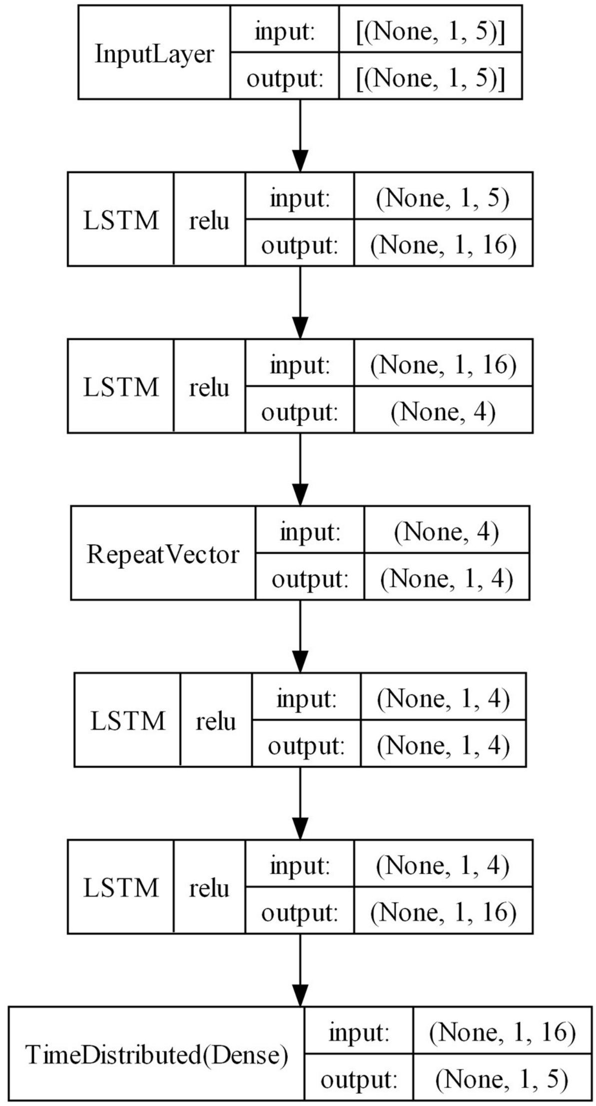

Fig. 5—Network based on LSTM architecture.

in which k controls the tolerance interval around the mean and can be set for each problem. In this study, $k = 3$ was adopted in accordance with the database used, comprising about 99.7% of the normal data saved. The threshold for occurrence of an anomalous event is based on this tolerance. If a new data SMAPE value is under this tolerance, the new instance is a normal datum. If SMAPE value exceeds tol_{AI} value, it can be associated with risk levels defined by the operator, supporting the real-time decision-making process. For the purpose of the system presented herein, values exceeding by 30%, 70%, and 100% represent low-, medium-, and high-risk anomalies, respectively. **Fig. 6** illustrates the process flow diagram triggered on a fixed time interval for anomaly detection using the developed LSTM-based autoencoder.

By using both literature (Vargas et al. 2019) and industry data, the autoencoder network was tested to verify its robustness when exposed to time series containing normal, transitional, and anomalous points. It was able to correctly classify 98–99% of the anomalous data, indicating that the autoencoder network is capable of accurately detecting undesirable events in oil production. **Fig. 7** shows an example of an abnormal state using real data from a subsea well, in which can be seen the pressure values gathered from PDG, PT, and TPT in the upper graph, the temperature values from PDG and TPT in the central graph, and the anomaly score value calculated by the neural network in the lower graph, compared with the tolerance line (threshold).

As previously stated, three parameters are obtained after the training—(1) data normalization parameters, (2) weights of neural networks, and (3) anomaly tolerance calculated during training. The first two are configuration parameters relevant to the transformation of input variables and network coefficients. As the monitoring is occurring in real time, and the conditions of the oilwell operation are constantly changing, new training is performed every time 500 new instances are saved. The number of instances required for new training was determined via hyperparameter testing on real data, aiming to have a sufficient amount of data that minimizes the hypersensitivity of the results without penalizing computational resources or causing overfitting. It should be noted that the weights, normalization parameters, and tolerance depend on the valve schematics. If the operator changes the valve configuration, a new network is trained specifically for that new scheme. In other words, each valve configuration has its own deep network associated with it, which evolves over time by the acquisition of new data. It also allows the simulation of a lookahead scenario, to predict the system response to a specified action, such as a valve maneuver.

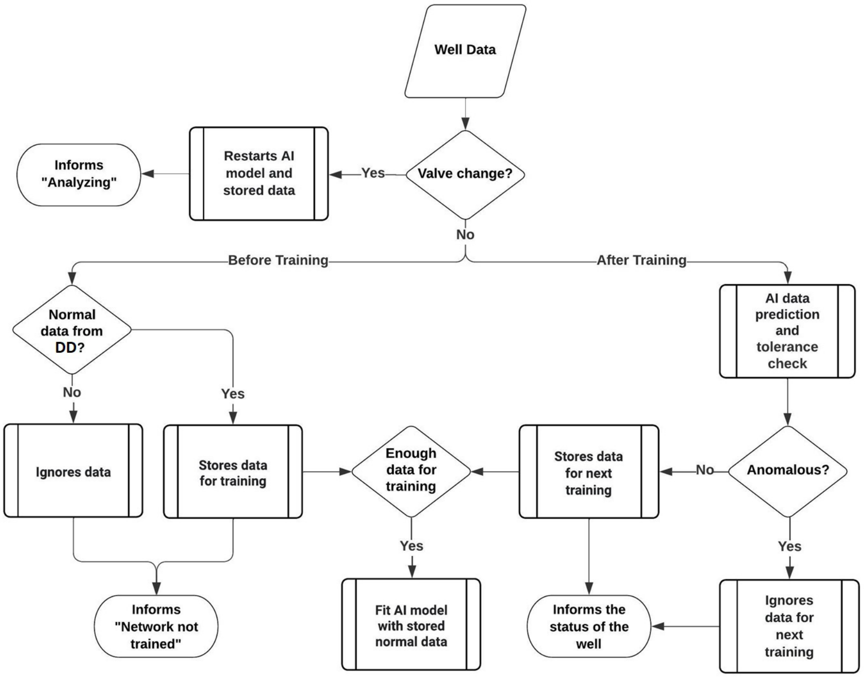

Fig. 6—Process flow diagram for anomaly detection using the developed LSTM-based network.

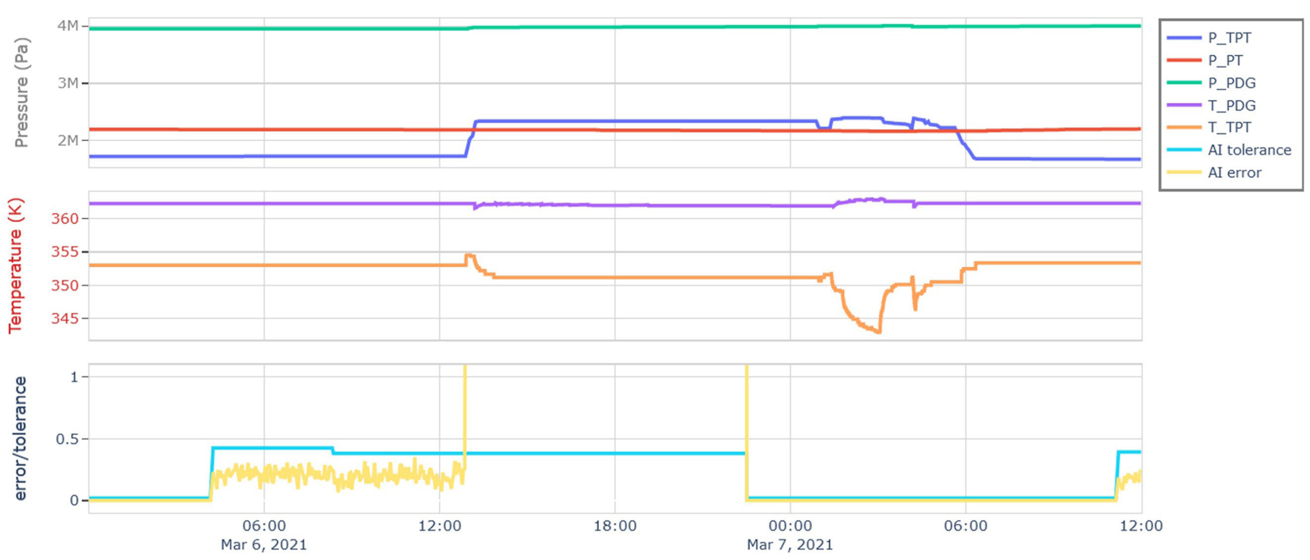

Fig. 7—Illustrative example of anomaly detection using AI.

Before training a new network, the data is scaled to be between 0 and 1 using a min-max scaler, to ensure that each variable has the same order of magnitude, leading to more stable and expedited training. The procedure is described in Eq. 3, in which x and x' refer to the original data point and its scaled value, respectively:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}. \tag{3}$$

This step is crucial to achieve accurate optimization results from the network. However, in some cases, the sensors installed in the wells may transmit data with low variation due to physical conditions or resolution limitations. This scenario can cause the network to become more sensitive, resulting in maximum and minimum values that are close together, so even minor variations may be unproperly classified as an anomalous event. To avoid this excessive sensitivity of the network from occasional low variation of the signal, randomly generated noise from a uniform distribution is added before scaling, as depicted in Eq. 4:

$$y = x' + \varepsilon. \tag{4}$$

The noise level ε is set between 0% and 0.5% of the average of the received data. This procedure not only enriches the data set by introducing variance but also ensures that the model is better equipped to distinguish between significant patterns and minor deviations, reducing the propensity for false positives.

Metrics Applied for Performance Evaluation. Classification metrics are highlighted and used as a benchmark to evaluate the performance of classification algorithms. Therefore, a set of metrics was used to compare and evaluate the algorithms. Accuracy (ACC) considers all normal and fault samples. ACC can be calculated using Eq. 5:

Classification metrics are highlighted and used as a benchmark to evaluate the performance of classification algorithms. Therefore, a set of metrics was used to compare and evaluate the algorithms. Accuracy (ACC) considers all normal and fault samples. ACC can be calculated using Eq. 5:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{n_{\mathrm{total}}}, \tag{5}$$

where TP is the number of true positives, TN is the number of true negatives, and n_{total} is the total number of samples. When dealing with unbalanced data, accuracy (ACC) is not a suitable measure. Instead, it is recommended to use balanced accuracy (ACC_b) as a more appropriate metric. The calculation for ACC_b can be estimated using Eq. 6:

$$\mathrm{ACC}_b = \frac{\mathrm{REC} + \mathrm{SP}}{2}, \tag{6}$$

where Recall (REC) indicates the proportion of anomalous data that is correctly detected from all anomalies. Typically, in industrial applications, REC is a prominent metric, as false negatives lead to much more harmful results than false alarms or false positives. The REC is given by Eq. 7:

$$\mathrm{REC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{7}$$

in which TP is the number of true positives, TN is the number of true negatives, and FN is the number of false negatives. The specificity (SP) estimates the ability of the algorithm to predict true negatives over false positives and can be calculated by Eq. 8:

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{8}$$

where FP is the number of false positives.

The F1-SCORE consists of an important metric, especially in problems containing imbalanced data. It is defined as a harmonic mean between recall (REC) and precision (PR) and can be estimated in Eq. 9:

$$\text{F1-SCORE} = \frac{2\left(\mathrm{PR} \times \mathrm{REC}\right)}{\mathrm{PR} + \mathrm{REC}}. \tag{9}$$

Precision (PR) evaluates how many of the predicted positive instances were actually true positives. It is defined by the ratio of true positives to the sum of true positives and false positives, as defined in Eq. 10:

$$\mathrm{PR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{10}$$

Additionally, the instances identification percentage (IIP) was calculated when the model was tested against instances of the same class of the target class. This metric is the percentage of the total instances of class in the test set for which the classifier successfully identified both phases—the normal and fault phases. A value of 1.0 indicates that the model correctly identifies both normal and fault occurrences in all tested instances.

Comparison with Other Methods

The proposed methodology has been validated against other classification techniques, addressing several data sets related to anomalies in subsea production/injection wells. To illustrate this process, the results of this study were compared with those of previous works on the same type of anomaly using RF (Marins et al. 2021), DT (Turan and Jaschke 2021), and LTSM autoencoder (Machado et al. 2022). This comparison refers to spurious DHSV closure events in 22 wells recorded in a public database (Vargas et al. 2019). The same type of anomaly is addressed in the "Case Studies" section for Well 1 and Well 3.

Case Studies

The proposed methodology was integrated into a real-time monitoring system of offshore wells, gathering data from sensors and the valve status of subsea production/injection wells. It has been implemented in more than 20 FPSOs, monitoring more than 250 wells in Santos and Campos basins, eastern Brazil, all equipped with multiplexed wet Christmas trees. An advantage of this system lies in its comprehensiveness in identifying anomalies related to well integrity for hundreds of subsea wells. These anomalies would be impossible for human operators to track. Some use cases are presented to illustrate the system's efficacy compared with human post-classification (class). Additionally, a web-based prototype was developed to provide a user-friendly experience for monitoring multiple wells simultaneously.

Three case studies are presented to illustrate the efficacy of the system in real monitoring situations:

Well 1 is an oil production well in the presalt area in Santos Basin, ultradeep water scenario, water depth 2041 m, drilled in five phases, directional trajectory, and final depth of 5870 m, equipped with downhole interval completion valves, and producing in three different zones at 5432–5774 m.

Well 2 is an oil production well in the presalt area in Santos Basin, water depth of 2125 m, drilled in four phases with directional trajectory and final depth of 5552 m, equipped with downhole interval completion valves, and producing in two different zones at 5137–5418 m.

Well 3 is a water-alternating-gas injection well in the presalt area in Santos Basin, in water depth of 1946 m, drilled in four phases, vertical trajectory and final depth of 6197 m, equipped with downhole interval completion valves, and injecting in three different zones at 5437–5738 m.

Results and Discussion

As mentioned above, the proposed methodology was applied to analyze 22 events recorded in a public database to demonstrate its validity. It was also applied in three case studies selected from the real-time monitoring system, to which the methodology was integrated, aiming to illustrate its efficacy. The results of these analyses are presented and discussed in this section.

Comparison with Other Methods. Table 1 presents the summarized results for the four classifiers applied to spurious DHSV closure data. RF was used for the results of Marins et al. (2021), DT for the model presented in Turan and Jaschke (2021), and LTSM refers to the results obtained by Machado et al. (2022) using LSTM autoencoder. Finally, DD + LSTM stands for the present study, and the comparison is performed in terms of ACC, F1-SCORE, and IIP metrics.

Classifiers	Accuracy (ACC)	F1-SCORE	IIP
RF	0.8708	–	0.58
DT	0.6000	0.4900	–
LSTM	0.9992	0.9360	1.0
DD + LSTM	0.9894	0.9917	1.0

Table 1—Comparison of metrics between literature and the proposed method for identifying spurious closure events of DHSV from a public database (Vargas et al. 2019).

From **Table 1**, the results obtained are in line with other works in the literature. The accuracy achieved is around 98.9%. On the same data set, Machado et al. (2022) reported 99.9% accuracy by using the LSTM autoencoder, Marins et al. (2021) presented 87.1% by using RF, and Turan and Jaschke (2021) reported 60% with the DT method. Looking at F1-SCORE values, DD + LSTM performs the best, followed by the LSTM model, both of which are significantly superior to the DT. The IIP values indicate that both DD + LSTM and LSTM models successfully identified both normal and anomaly instances, in contrast to the RF results for this metric.

The coupled method has demonstrated superior performance compared with using the DD alone. To illustrate this, the results obtained by using the DD model are compared with the ones provided by the comprehensive model (DD +LSTM). **Table 2** presents the average values for accuracy (ACC), balanced accuracy (ACC_b), and F1-SCORE, allowing for a direct comparison between the models developed herein.

	ACC	ACC_b	F1-SCORE	IIP
DD	0.9727	0.9615	0.9758	1.0
DD + LSTM	0.9894	0.9771	0.9917	1.0

Table 2—Metrics obtained for the DD model and the complete model (DD +LSTM) for identifying spurious closure events of DHSV from a public database (Vargas et al. 2019).

Case Studies. The dual system proposed was applied to the three case studies, and the results were compared with human post-classification (class), as presented in **Figs. 8 through 11.** It can be noticed that the occurrence of anomalies happens abruptly in all case studies. In these figures, the bottom graphics represent the statuses of the DD and LSTM methods, which can be labeled as follows: Analyzing, for both methods, which is triggered when there is a change of a valve status (e.g., M1 has closed), that will cause temporary fluctuations in the pressures and temperatures measured by the sensors, so the system awaits some time for these readings to stabilize before resume monitoring; Network not trained, exclusive to LSTM, reported just after Analyzing while there are not enough data classified by the DD as normal for the LSTM to train; and Normal/Anomalous, for both methods, regarding anomaly detection; and some others exclusive to DD, such as DHSV/M1 closed, reported when according to sensors those valves should be open, but the DD detects that they might actually be closed.

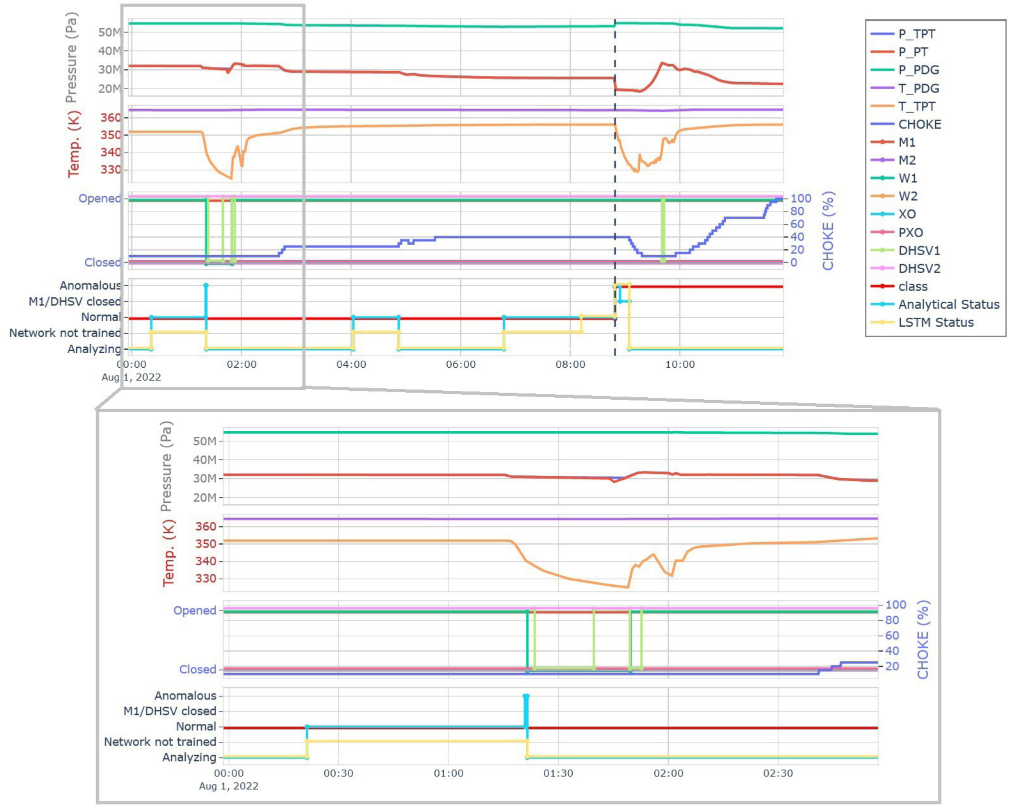

Fig. 8—Results for Well 1 for the first anomalous event detected—PDG, PT, and TPT sensors, valve status, and comparison between model classification (LSTM Status and Analytical Status) and fault classification (class).

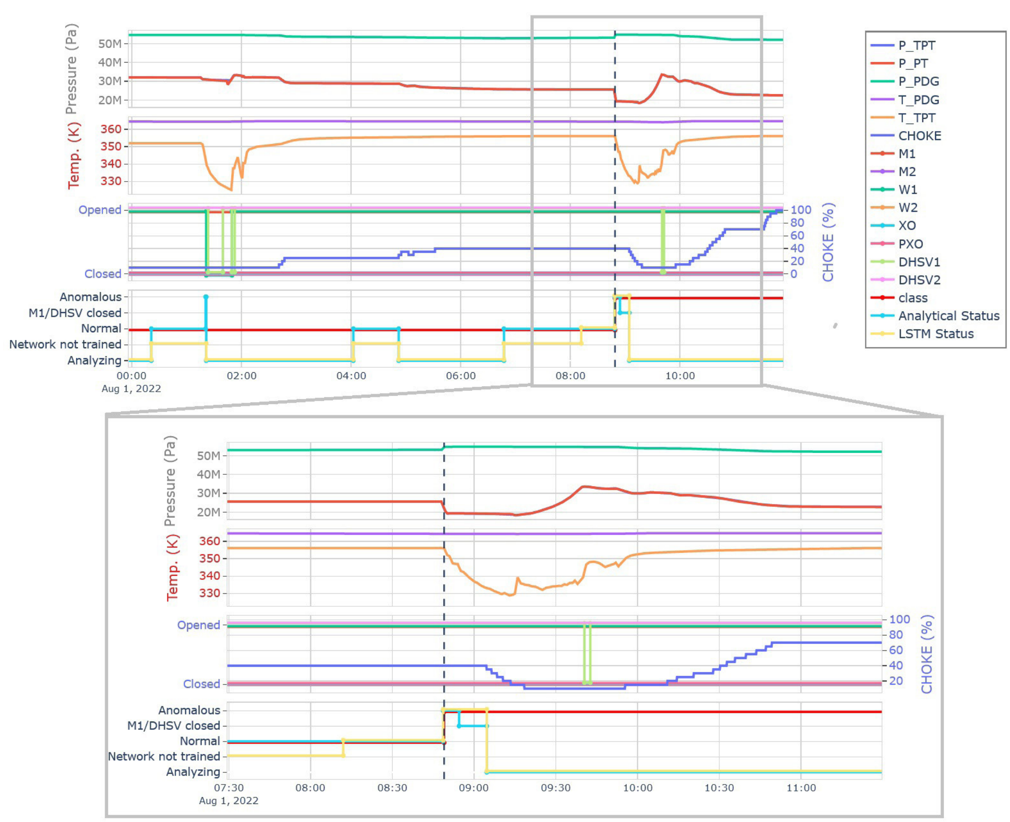

Fig. 9—Results for Well 1 for the second anomalous event detected—PDG, PT, and TPT sensors, valve status, and comparison between model classification (LSTM Status and Analytical Status) and fault classification (class).

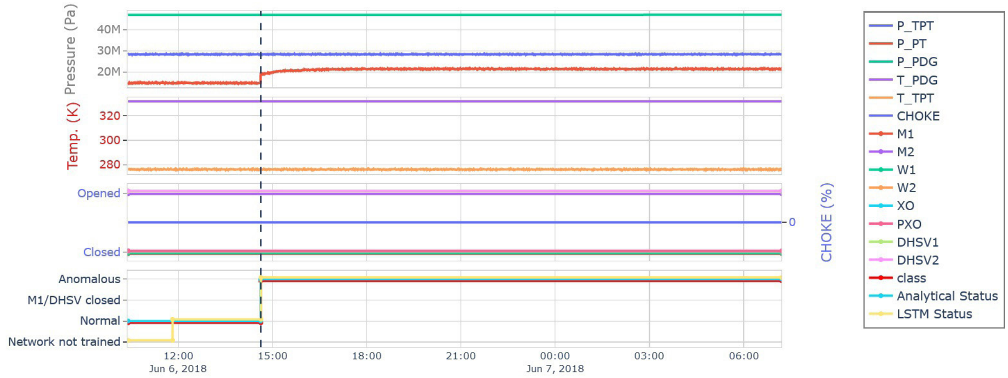

Fig. 10—Results for Well 2—PDG, PT, and TPT sensors, valve status, and comparison between model classification (LSTM status and analytical status) and fault classification (class).

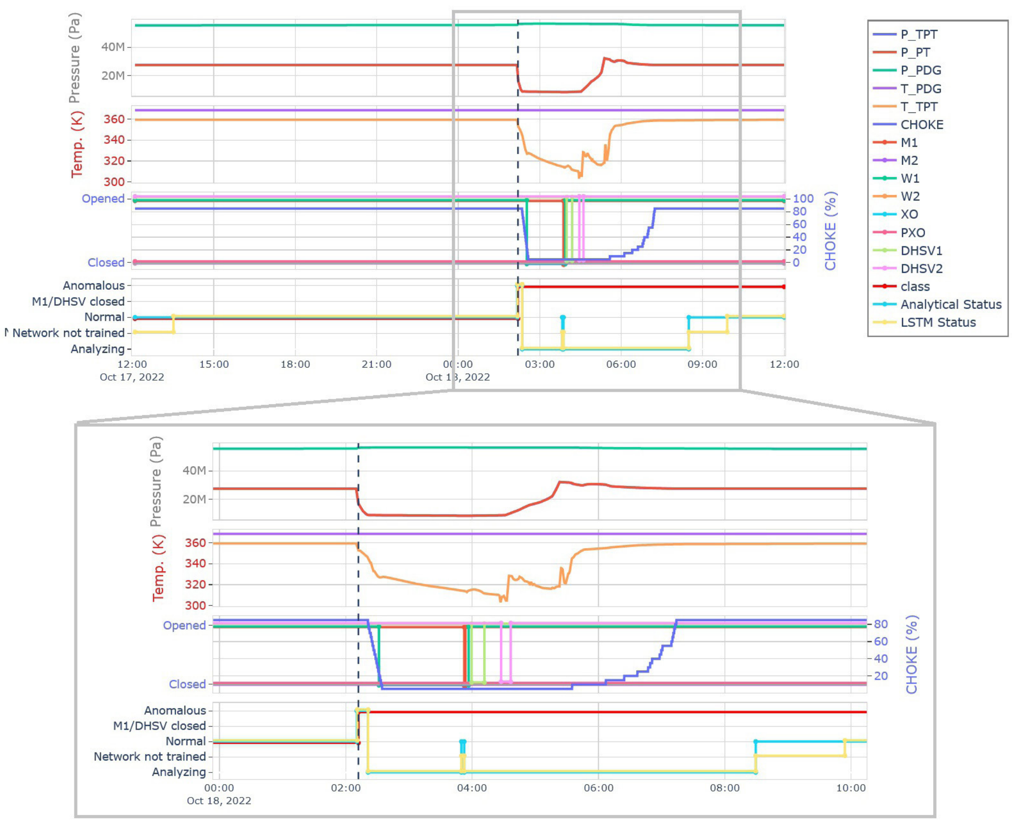

Fig. 11—Results for Well 3—PDG and TPT sensors, valve status, and comparison between model classification (LSTM status and analytical Status) and fault classification (class).

The analyzed data from Well 1 are presented in **Fig. 8** and show data from the PDG, PT, TPT, and valve status sensors along with the results of analyses and classifications for a specific time window on 1 August 2022. All the valves have a binary open or closed status except for the choke, which reports a partial opening indication. Two anomalous events were identified in the period. The first at 1:16 a.m. (curve analytical status) related to a spurious closure of the DHSV, corrected by the operator in the sequence with valve cycling performed from 1:39 a.m. to 1:52 a.m. Only the analytical approach captured this anomaly, whereas the AI did not have time to identify it because the system went into analysis mode due to the change in parameters. A few hours later **(Fig. 9)**, a second event at 8:48 a.m. also occurred, being identified by the model (curves LSTM status and analytical status) and promptly corrected by the operator. In this case, the analytical model correctly classified, after detecting the anomaly, as spurious closure of the DHSV or the M1 and the LSTM model correctly identifies the anomalous event.

The following example refers to Well 2, In this well, the developed system was applied in a retroanalysis to identify an anomaly event at 2:37 p.m. on 6 June 2018 (curves LSTM status and analytical status). In the field, the problem was only detected sometime later in a light workover intervention; the cause was the pressure communication by the gas lift valve with the subsea well closed. **Fig. 10** presents the data from the PDG, PT, TPT, and valve status sensors along with the results of the analyses and classifications. It is important to mention that, in this well, if the developed methodology had already been implemented at that time, the identification and correction of the problem would have been anticipated, avoiding NPT, so bringing enormous gains in oil production.

Fig. 11 presents data and results for Well 3, on 18 October 2022. The dual model identified the anomaly at 2:10 a.m. (curves LSTM status and analytical status) and correctly classified it in the sequence as spurious closure of DHSV or M1. It is important to note that the operator then closed the well and proceeded with the valve cycling and reopening procedure, resuming production in safe condition. After reopening well at 7:14 a.m., the system restarts to analyze and returns after some time to normal condition at 8:29 a.m.

Table 3 presents the SMAPE values, balanced accuracy (ACC_b), and F1-SCORE values calculated for the cases studied when compared with the human post-classification (class).

	SMAPE	ACC_b	F1-SCORE
Well 1 - first event	39.148	*	*
Well 1 - second event	48.312	0.9987	0.9987
Well 2	33.136	0.9998	0.9998
Well 3	23.390	0.9968	0.9991

*Human post-classification did not identify the anomalous event.

Table 3—Comparison of metrics between literature methods and the proposed method for the three case studies.

The dual model consistently exhibits superior performance across all metrics when contrasted with the standalone DD model. This is further evident in the detection results presented in **Table 4,** where ACC_b and F1-SCORE are compared for the case studies. Although the DD model on its own already produces impressive results, the combined model surpasses it.

	DD		DD + LSTM	
	ACC_b	F1-SCORE	ACC_b	F1-SCORE
Well 1 - second event	0.9800	0.9306	0.9886	0.9373
Well 2	0.9990	0.9998	0.9998	0.9998
Well 3	0.9988	0.9983	0.9995	0.9993

Table 4—Metrics obtained for the DD model and the complete model (DD + LSTM) for the three case studies.

Some Comments on the Proposed System. The proposed system performs well on the task of anomaly detection, presenting detection rates exceeding 90% in the real field production scenarios studied. Changes in well and reservoir conditions over time impose concept and data drift in time series, which are challenging factors for machine learning models. The dual approach enables the handling of these aspects, albeit generating the need for recurrent training of LSTM autoencoder, which can result in increased computational cost. The DD part serves the specific and valuable purpose of describing the normal state of the well under different valve states. However, due to its analytical basis, it exhibits lower sensitivity compared with LSTM, resulting in a lower detection rate.

DD is primarily used to assess the system's status under different valve schemes, relying on predefined rules and up to a predetermined threshold of 5%. However, in some cases, minor variations can result in false positives, making it a complementary method to LSTM. In contrast, LSTM exhibits dynamic sensitivity that adapts with each training, enabling it to detect subtle changes that the analytical method may occasionally miss.

Prototype Development. Aiming to provide an agile and accurate monitoring experience, a web-based computational system has been developed, allowing for handling of multiple wells simultaneously. From the main dashboard, the operator can set a specific well, whose graphical interface is presented in **Fig. 12.** The center of the screen displays two graphs on the left that represent the absolute values of the sensors, as well as two first graphs on the right that display the relative error of the moving average calculated from previous data. The lower right graph shows the reconstructed error and tolerance from the AI approach. On the sides, the latest absolute values received from the sensors are displayed on the left, while the latest calculated relative errors are displayed on the right. At the top of the screen, the well status evaluated by the analytical approach and the neural network is shown, along with some real-time monitoring control buttons.

In addition to searching for anomalies, the system has some functionalities that help prevent false positives. One of them is the detection of frozen sensors, which evaluates the repetition of values from the sensors for a certain period of time. If detected, the frozen sensors are reported at the system output. Another functionality is the identification of invalid data, so that inconsistent data, such as nonnumeric or null values for sensors, are not included in the analysis. Changes in the valve status can significantly alter the well behavior. Therefore, the system has an analyzing mode whose objective is to provide adequate time for pressure and temperature levels to stabilize, thus ensuring that planned changes are not mistaken for anomalous events. With these functionalities, the system becomes more robust and capable of providing more security during monitoring.

Conclusions

In this work, we describe a system for real-time monitoring and anomaly detection using production sensor data from oil wells. The system utilizes a combination of deep learning autoencoder and a rule-based analytic approach for detecting unexpected events, with the goal of improving operational safety and reducing costs associated with NPT and failure repair.

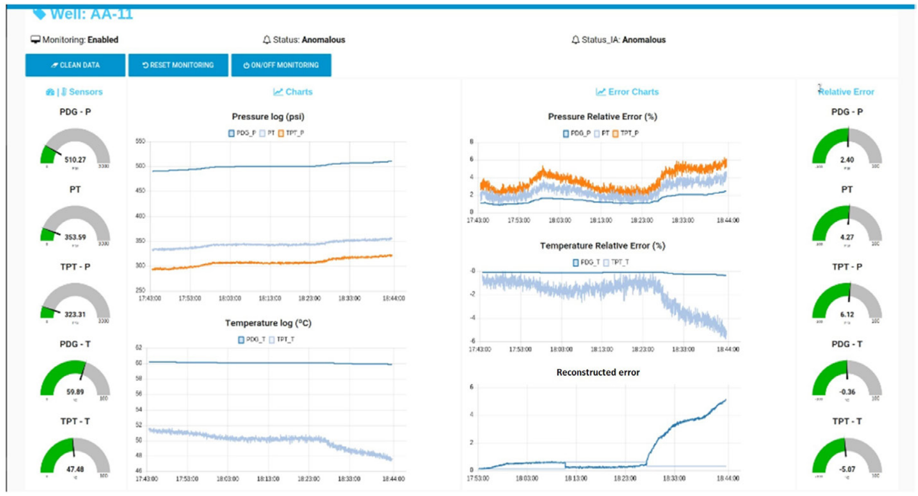

Fig. 12—Illustrative example of the developed prototype: Parameters and system response during a tubing to annular communication anomalous event.

Initial studies were conducted to determine the behavior and correlations of pressure and temperature values for the most common combinations of well valve states. The proposed methodology uses pressure and temperature sensor data to classify the well status via a DD, which is then used to train autoencoders based on LSTM deep networks devoted to anomaly detection. This coupling enables the deep neural network to evolve constantly through the normal data collected by the analytical method. It makes the system versatile enough to adapt itself to variations of the valve scheme while also allowing new anomalies to be added to the rule-based analytical approach's cataloged portfolio.

A comparison with other approaches using the same public data set, focusing on spurious DHSV closure events, is provided. Evaluating metrics such as accuracy, F1-SCORE, and IIP indicates that the proposed system either performs comparably or outperforms other machine Learning techniques.

The developed system exhibits high accuracy, with true-positive detection rates exceeding 90% in the early stages of anomalies identified in both simulated and actual well production scenarios. From these latter, it can be highlighted the enhancement of the detection process, typically carried out by humans. For the case studies presented, all three wells had a short transitional phase, and the model performed well with a balanced accuracy higher than 98.0%.

The system is implemented in more than 20 FPSOs, monitoring more than 250 production/injection subsea wells, and can be applied in both real-time operation and testing scenarios. The continuous acquisition of data through sensors combined with the adaptability of the detection model ensures robustness over time, along with the changes in well, reservoir, and operation conditions. It is worth mentioning the limited computational infrastructure available at rigsite, which requires expedited models to provide a quick response to the data flow. By observing the performance of the proposed system on the massive simultaneous monitoring, it has consistently delivered a prompt response in online well integrity monitoring.

In summary, the proposed system for online monitoring and anomaly detection has been validated successfully and applied to the analysis of real-time data. It exhibits high accuracy rates in detecting anomalies, making it a valuable tool for supporting the decision-making process in oilwell monitoring.

However, there is still room for future research such as a rigorous study on how accurately the proposed method can perform early-stage anomaly detection and to address persistent challenges related, for example, to valve performance and tightness. While this work focuses on field application, future research could explore the use of other deep learning architectures, such as gated recurrent unitss or transformer neural networks, which are known for their efficiency in processing sequential data.

Nomenclature

DHSV = down hole safety valve
ICV = interval completion valve
M1 = master valve 1 in wet Christmas tree
M2 = master valve 2 in wet Christmas tree
W1 = wing valve 1 in wet Christmas tree
W2 = wing valve 2 in wet Christmas tree
PXO = pig cross-over valve in wet Christmas tree
XO = cross-over valve in wet Christmas tree

References

Al-Hajri, N. M., Al-Ghamdi, A., Tariq, Z. et al. 2020. Scale-Prediction/Inhibition Design Using Machine-Learning Techniques and Probabilistic Approach. *SPE Prod & Oper* **35** (4): 0987–1009. https://doi.org/10.2118/198646-PA.

Alharbi, B., Liang, Z., Aljindan, J. M. et al. 2022. Explainable and Interpretable Anomaly Detection Models for Production Data. *SPE J.* **27** (01): 349–363. https://doi.org/10.2118/208586-PA.

Anifowose, F., Abdulraheem, A., and Al-Shuhail, A. 2019. A Parametric Study of Machine Learning Techniques in Petroleum Reservoir Permeability Prediction by Integrating Seismic Attributes and Wireline Data. *J Pet Sci Eng* **176**: 762–774. https://doi.org/10.1016/j.petrol.2019.01.110.

Cheng, J., Mao, D., Salamah, M. et al. 2022. Scale Buildup Detection and Characterization in Production Wells by Deep Learning Methods. *SPE Prod & Oper* **37** (04): 616–631. https://doi.org/10.2118/205988-PA.

Chollet, Fet al. 2015. *Keras*. https://github.com/fchollet/keras.

D'Almeida, A. L., Bergiante, N. C. R., de Souza Ferreira, G. et al. 2022. Digital Transformation: A Review on Artificial Intelligence Techniques in Drilling and Production Applications. *Int J Adv Manuf Technol* **119** (9–10): 5553–5582. https://doi.org/10.1007/s00170-021-08631-w.

Figueirêdo, I. S., Carvalho, T. F., Silva, W. J. D. et al. 2021. Detecting Interesting and Anomalous Patterns In Multivariate Time-Series Data in an Offshore Platform Using Unsupervised Learning. Paper presented at the Offshore Technology Conference, Virtual and Houston, Texas, USA, 16–19 August. https://doi.org/10.4043/31297-MS.

Gjelsvik, E. L., Fossen, M., and Tøndel, K. 2023. Current Overview and Way Forward for the Use of Machine Learning in the Field of Petroleum Gas Hydrates. *Fuel* **334**: 126696. https://doi.org/10.1016/j.fuel.2022.126696.

Gurina, E., Klyuchnikov, N., Zaytsev, A. et al. 2020. Application of Machine Learning to Accidents Detection at Directional Drilling. *J Pet Sci Eng* **184**: 106519. https://doi.org/10.1016/j.petrol.2019.106519.

Hüffner, L. N., Trierweiler, J. O., and Farenzena, M. 2019. Are Complex Black-Box Models for Permanent Downhole Gauge Pressure Estimation Necessary? *J Pet Sci Eng* **173**: 715–732. https://doi.org/10.1016/j.petrol.2018.10.047.

Kingma, D. P. and Ba, J. 2014. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 (preprint; last revised 30 January 2017). https://doi.org/10.48550/arXiv.1412.6980.

Larzalere, B. 2019. *LSTM Autoencoder for Anomaly Detection*. https://towardsdatascience.com/lstm-autoencoder-for-anomaly-detection-e1f4f2ee7ccf (accessed 30 October 2019).

Li, D., Chen, D., Goh, J. et al. 2018. Anomaly Detection with Generative Adversarial Networks for Multivariate Time Series. arXiv:1809.04758 (preprint; last revised 15 January 2019). https://doi.org/10.48550/arXiv.1809.04758.

Ma, J. and Perkins, S. 2003. Time-Series Novelty Detection Using One-Class Support Vector Machines. Paper presented at the 2003 International Joint Conference on Neural Networks, 1741–1745, Portland, Oregon, USA. https://doi.org/10.1109/IJCNN.2003.1223670.

Machado, A. P. F., Vargas, R. E. V., Ciarelli, P. M. et al. 2022. Improving Performance of One-Class Classifiers Applied to Anomaly Detection in Oil Wells. *J Pet Sci Eng* **218**: 110983. https://doi.org/10.1016/j.petrol.2022.110983.

Malhotra, P., Vig, L., Shroff, G. et al. 2015. Long Short-Term Memory Networks for Anomaly Detection in Time Series. Paper presented at the 23rd European Symposium on Artificial Neural Networks, ESANN 2015, Bruges, Belgium, 22–24 April. https://www.esann.org/sites/default/files/proceedings/legacy/es2015-56.pdf.

Marchi, E., Vesperini, F., Eyben, F. et al. 2015. A Novel Approach for Automatic Acoustic Novelty Detection Using A Denoising Autoencoder with Bidirectional LSTM Neural Networks. Paper presented at the ICASSP 2015 - 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1996–2000, South Brisbane, Queensland, Australia. https://doi.org/10.1109/ICASSP.2015.7178320.

Marins, M. A. 2018. *Machine Learning Techniques Applied to Hydrate Failure Detection on Production Lines*. Master's Thesis, Federal University of Rio de Janeiro.

Marins, M. A., Barros, B. D., Santos, I. H. et al. 2021. Fault Detection and Classification in Oil Wells and Production/Service Lines Using Random Forest. *J Pet Sci Eng* **197**: 107879. https://doi.org/10.1016/j.petrol.2020.107879.

Mattia, F. D., Galeone, P., Simoni, M. D. et al. 2019. A Survey on GANs for Anomaly Detection. arXiv:1906.11632 (preprint; last revised 14 September 2021). https://doi.org/10.48550/arXiv.1906.11632.

Monday, C. U. and Odutola, T. O. 2021. Application of Machine Learning in Gas-Hydrate Formation and Trendline Prediction. Paper presented at the SPE Symposium: Artificial Intelligence - Towards a Resilient and Efficient Energy Industry, Virtual, 18–19 October. https://doi.org/10.2118/208653-MS.

Sabokrou, M., Khalooei, M., Fathy, M. et al. 2018. Adversarially Learned One-Class Classifier for Novelty Detection. Paper presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, Utah, USA, 18–23 June. https://doi.org/10.1109/CVPR.2018.00356.

Shawe-Taylor, J. and Zlicar, B. 2015. Novelty Detection with One-Class Support Vector Machines. In *Advances in Statistical Models for Data Analysis*, eds. I. Morlini, T. Minerva, and M. Vichi, 231–257. New York, USA: Springer International Publishing. https://doi.org/10.1007/978-3-319-17377-1.

Tang, J., Fan, B., Xiao, L. et al. 2021. A New Ensemble Machine-Learning Framework for Searching Sweet Spots in Shale Reservoirs. *SPE J.* **26** (01): 482–497. https://doi.org/10.2118/204224-PA.

Turan, E. M. and Jaschke, J. 2021. Classification of Undesirable Events in Oil Well Operation. Paper presented at the 2021 23rd International Conference on Process Control (PC), Strbske Pleso, Slovakia, 1–4 June. https://doi.org/10.1109/PC52310.2021.9447527.

Vargas, R. E. V., Munaro, C. J., Ciarelli, P. M. et al. 2019. A Realistic and Public Dataset with Rare Undesirable Real Events in Oil Wells. *J Pet Sci Eng* **181**: 106223. https://doi.org/10.1016/j.petrol.2019.106223.

Vávra, J. and Hromada, M. 2019. Optimization of the Novelty Detection Model Based on LSTM Autoencoder for ICS Environment. In *Intelligent Systems Applications in Software Engineering. CoMeSySo 2019 2019. Advances in Intelligent Systems and Computing*, eds. R. Silhavy, P. Silhavy, and Z. Prokopova, Vol. 1046, 306–319. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-30329-7.

Venkatasubramanian, V. 2003. A Review of Process Fault Detection and Diagnosis Part i: Quantitative Model-Based Methods. *Comput Chem Eng* **27** (3): 293–311. https://doi.org/10.1016/S0098-1354(02)00160-6.

Wang, S. and Chen, S. 2019. Application of the Long Short-Term Memory Networks for Well-Testing Data Interpretation in Tight Reservoirs. *J Pet Sci Eng* **183**: 106391. https://doi.org/10.1016/j.petrol.2019.106391.

Yousefzadeh, R., Bemani, A., Kazemi, A. et al. 2023. An Insight into the Prediction of Scale Precipitation in Harsh Conditions Using Different Machine Learning Algorithms. *SPE Prod & Oper* **38** (02): 286–304. https://doi.org/10.2118/212846-PA.

Zhao, X., Jin, F., Liu, X. et al. 2022. Numerical Study of Fracture Dynamics in Different Shale Fabric Facies by Integrating Machine Learning and 3-D Lattice Method: A Case from Cangdong Sag, Bohai Bay Basin, China. *J Pet Sci Eng* **218**: 110861. https://doi.org/10.1016/j.petrol.2022.110861.

Zhong, Z., Sun, A. Y., Wang, Y. et al. 2020. Predicting Field Production Rates for Waterflooding Using a Machine Learning-Based Proxy Model. *J Pet Sci Eng* **194**: 107574. https://doi.org/10.1016/j.petrol.2020.107574.