1016

MACHINE LEARNING AND DEEP LEARNING IN OIL AND GAS INDUSTRY: A REVIEW OF APPLICATIONS, OPPORTUNITIES AND CHALLENGES

Tejas Balasaheb Sabale[1], Syed Aaquib Hussain[2], Mohd Zuhair[3], Mohammad Saud Afzal[1], Arnab Ghosh[2]

[1]Department of Civil Engineering, Indian Institute of Technology, Kharagpur, India
[2]SLB, Mumbai, India
[3]Nirma University, Ahmedabad, India

ABSTRACT

Oil and natural gas accounts for more than 57% of energy consumption, and the worldwide utilization is increasing at the rate of 1.8% for oil and 0.9% for gas. This paper analyzes and draws insights on machine learning techniques that have been implemented in the oil and gas industry at different levels across the entire value chain. This paper aims to provide a comprehensive state of art applications of Machine Learning (ML) and Deep Learning (DL) in the Oil and Gas Industry. The literature study is conducted to investigate the applications of ML and DL in different process chains (upstream, midstream, and downstream). The paper also attempts to build a taxonomy for the different processes of the hydrocarbon's exploration and production. It is a study of different ML processes involved in different sectors of the industry and how it has mitigated challenges to achieve a step change in operations in the entire oil and gas value chain.

INTRODUCTION

The global economy heavily relies on the oil and gas industry (OGI) to fulfill the energy needs to compete for industrial development.. The reduction of utilization of oil and gas is the common agenda globally by promoting different renewable sources of energy, still oil and gas will constitute to more than 50% of the global energy requirements till the year 2040 [1].

Apart from meeting the energy requirements, the oil and gas and its by-products are raw products for petro-chemical industries such as chemicals, drugs, fertilizers, pesticides etc. The increasing demand of the outputs from the OGI to integrate with the new technologies is rising faster than ever. Moreover, the pricing of oil and gas has a rippling impact which reduces overall economic growth, GDP, and increases inflation [2]–[4]. The increasing prices have often led to economic recessions in countries [5].

The volatile nature of the prices is impacting both the resource rich as well as import dependent countries. Therefore, the oil and gas companies are bringing in new technologies into their production line to reduce the losses and increase the efficiency. The technology will help the industry by creating network-based data communication to generate information for knowledge creation and to optimize the production cost [6]. In recent years, industry has adopted and developed new technologies for intelligent drilling, digital platforms, drilling robots etc. [7], [8].

The different processes of the industry generate an enormous amount of data, referred to as big Data, presented in Table 1. Analyzing and drawing insights from data could easily be done with the use of machine learning (ML) and deep learning (DL). The ML and DL techniques have been implemented in the industry at different levels for different sectors [8], [9]. The operations in the OGI are divided into three groups: upstream, downstream, and midstream. Leveraging ML and DL techniques in an efficient way can help in better decision-making across all sectors. This can ensure proper functioning of all the components of the systems to ensure early detection of faults [10].

Section	Data Generation
Drilling	0.3 GB/well/day
Seismic	1 GB/sq km
Pipeline Inspection	1.5 TB/600km
Vibration	7.5 GB/year/customer
Plant Operational	8 GB/year
Wireline	5 GB/well/day

Table 1. Different processes generating amount of data [10]

BACKGROUND

This section focuses on the basic sectors of the oil and gas industry, ML and DL.

A. Oil and Gas

The OGI can be divided into three key sectors upstream, midstream and downstream. The activities in OGI directly or indirectly impact everybody's life due to its usage in day-to-day life like transportation, electricity etc. [11].

Upstream: It is also called the E&P (Exploration and Production) sector that involves all the activities starting from exploring for a potential source of oil and gas to finally extracting it. It starts with a geological assessment for determining whether or not a given land or sea prospect is viable for drilling, followed by the use of geophysical methods to locate the presence of subsurface bodies. Gravity, magnetic, electromagnetic, and seismic methods are used for the same. Exploratory wells are then drilled to test the resource and the process of extracting the hydrocarbons is called production. The stream of petrophysics helps in the interpretation of the well log data and the stream of reservoir engineering is concerned with producing oil and gas reservoirs to maximize the extraction.

Midstream: The midstream activities mainly include transportation and storage. It links the upstream and downstream sectors. Oil and gas are transported via pipelines, tanker ships, trucks and railways. Finally, there are bulk terminals, refinery tanks, holding tanks, and underground reservoirs that store the oil and gas. Midstream assets and activities are found at any location where oil and gas is produced, transported, or sold.

Downstream: It involves refining crude oil, purifying the natural gas, marketing and commercial distribution of the finished products. Some of these products include LPG, petrol, diesel oil, natural gas, kerosene, lubricants, and gasoline.

B. ML and DL

The applications of ML and DL are already being deployed or envisioned in every sector of the industry. ML and DL, which is a subset of AI, helps systems to automatically learn and improve from experience, without being explicitly coded for each iteration. These techniques have huge potential to provide information from the heterogeneous data generated by different processes that helps in the decision-making process for different issues of OGI [9], [12].

Machine Learning: ML algorithms can process large amounts of data and help in extracting information using different models. The primary task of ML is to perform prediction, clustering, and decision making from the available input parameters [13], [14]. In OGI, it has the potential implementation in designing and developing the exploration plan, monitoring and diagnosing, predicting, forecasting and performing real-time optimization without much human intervention at a minimal cost [15], [16]. ML algorithms are mainly of three types: Supervised Learning, Unsupervised Learning, and Reinforcement Learning.

Supervised Learning (SL): In SL models are trained on 'labeled' data, implying that both input and output variables are known to the machine. The task is to find the best relationship between the input and output variables using different algorithms to predict the outcome correctly [17]. The model is initialized with control parameters and then the input data is fed into the model. Then the model adjusts its weights and parameters until the model has reached an acceptable accuracy of prediction, followed by the cross-validation process to check overfitting. After getting the satisfactory results from the validation process, the model can be used for testing new data points [18].

Unsupervised Learning: This type of learning also known as exploratory learning in the field of mathematics where the dataset used are 'unlabeled' [19]. Clear input and output are not known to the system and the task of the model is to draw inferences such as extract patterns, similarities, and differences among the data points without analyzing the output [20].

Reinforcement Learning: In this approach, a method of evaluating how good or bad the behavior of the model is used. The model rewards the desired behavior to encourage the agent and puts a penalty on the negative behavior [21]. This reward and penalty mechanism makes the agent look for maximizing the overall reward to attain the optimal solution of the problem [22].

Deep Learning: It is a subset of ML, uses neural networks which passes data through layers for learning. The word 'deep' in DL refers to the number of layers that are used in the neural network [23]. There is one main reason why DL is increasingly becoming popular as compared to traditional ML algorithms. After attaining a threshold, the traditional ML algorithms do not exhibit improvement even if we increase the amount of data used for training. But deep neural networks exhibit significant improvement as more and more data is supplied [24], [25]. It can handle colossal amounts of data and can perform complex tasks like well data analysis [26], image processing [25] etc.

C. Need of ML and DL in OGI

The oil and gas industry, similar to many other industries, is going under a digital transformation, often referred to as 'Oil and Gas 4.0'. The usage of real-time data and field actuating devices has increased exponentially in the upstream oil and gas industry, resulting in several digital oilfield installations. These have been shown to improve operational efficiency and increase production. Simultaneously, digital oilfield deployments have aided better and faster decision-making while lowering health, environmental, and safety threats. A digital oilfield's predictive analytics component is responsible for creating data-driven insight into the oilfield's current and future prospects. This component, which frequently incorporates ML algorithms, is critical to the effective implementation of digital oilfields [27].

APPLICATIONS

This section will discuss the various applications of ML and DL in the OGI, following a structured taxonomy as shown in Figure 1.

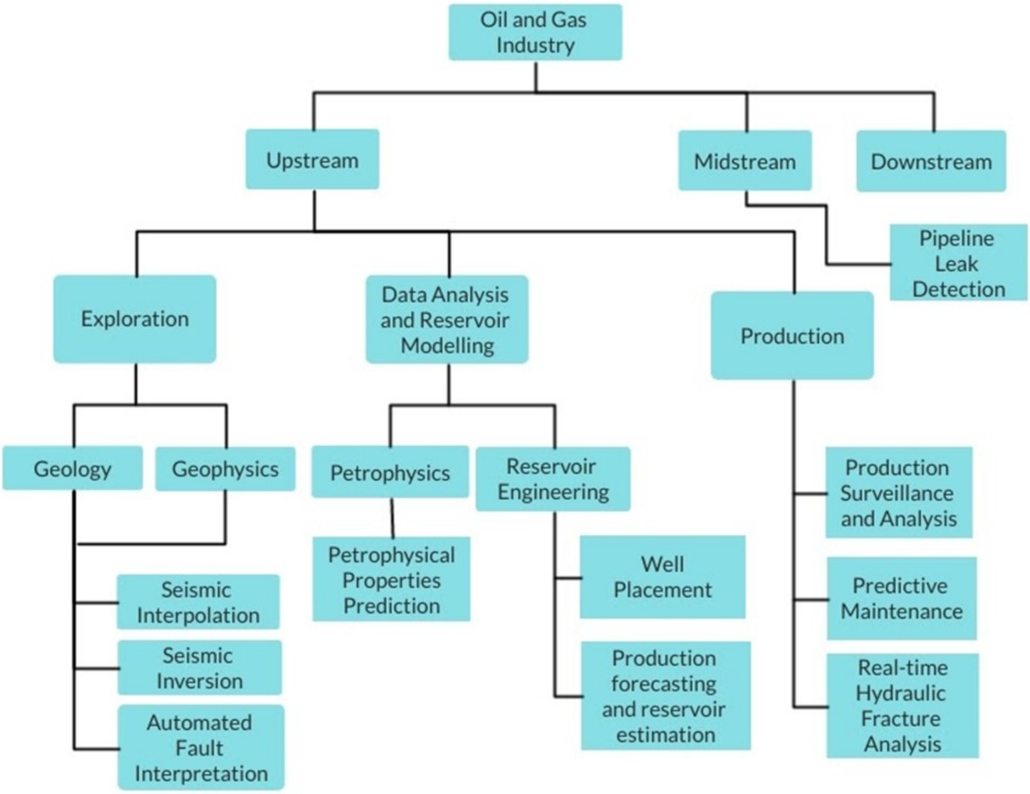

Figure 1. Applications of Machine Learning and Deep Learning in Oil and Gas Industry

1. Upstream:

The upstream oil and gas sector is under increasing pressure to reduce costs and increase efficiencies so it can remain competitive in an evolving landscape. In this paper, we discuss the ML and DL over widely incorporated different aspects including geology, geophysics, estimation of petrophysical properties, reservoir simulation, reserve estimation, reservoir performance, production, and forecasting, well placement and hydraulic fracturing optimization.

a. Exploration:

Drilling a well for extracting oil and gas is a big-budget costly process. If the drilled well turns out to be unproductive, it can be a huge waste of time, money, and resources. This is riddled with multiple risks as the explorationists need to identify subsurface prospects accurately for drilling and exploitation of hydrocarbons. And the only available information one has is from limited 2D seismic data. Hence, to increase the chances of drilling a productive well, it is of extreme importance that information of the Earth’s subsurface is collected and studied in a critical manner.

i. Geology

Oil and gas fields are usually located at places that face simultaneous occurrence of geologic features like oil and gas source rocks, migration, reservoir rocks, seals, and traps. Folded and faulted rock strata commonly form traps and accumulate fluids like petroleum and natural oils. Hence, geological assessment clearly plays a crucial role in identifying the best possible spots to drill. The interpretation of the geological data is left to experts of the domain, whose decisions are mostly based on personal skill, knowledge, and previous experiences. Therefore, it involves a lot of uncertainty.

ii. Geophysics

To obtain information about the subsurface, many geophysical methods like gravity, magnetic and seismic are implemented, seismic being the most common. The seismic waves are vibrations that travel through the ground. Based on the travel time of the reflected waves, various information is extracted. The seismic images provide detailed structure of the underlying surface. It shows faults, folds, orientation of the rock layers, shape, and size of these layers. These images can be 2D sections or 3D volumes (cubes) and are studied extensively to identify and pinpoint areas where oil and gas is most likely to occur. Moreover, the seismic images might have missing data in some portions due to various problems in data acquisition or processing issues. The further exploration processes heavily depend on the interpretation from these seismic records. Hence, it is important to have high-quality images for getting better insights.

1. Seismic Interpolation

Seismic processing techniques often assume data is regularly sampled in space, but acquisition methods, particularly in marine environments, rarely achieve this in practice. Data exhibiting large gaps and irregular sampling need interpolation and regularization prior to further processing. This being the critical component for exploration for data sources, constructing regularly shaped wavefields and filling in missing data is a dire need of the hour. The Generative Adversarial Networks (GAN) can be used for interpolation of these images. With the help of GANs, the bad traces can be reconstructed, and hence, better, and complete data can be used for analysis [28]. Furthermore, it can also be extended to improving the resolution of the images.

2. Seismic Inversion

The raw seismic data that is obtained needs to be processed before it can be used to extract information about the subsurface. This process of converting the seismic reflection data into properties of the rock is called inversion. The key to processing the seismic data is seismic velocity modeling. The task of inversion can be automated by training a model over seismic records mapped to their corresponding earth model and generating elastic profiles (P-wave velocity, S-wave velocity, density). From ML point of view, the task of inversion can be considered as a regression problem. The input for the DL model is seismic images. By training a CNN model on these images, we can simultaneously estimate all three elastic parameters (P-wave velocity, S-wave velocity, density). Training can be done on seismic records that are mapped to an earth model [29]. The results show a good match, but the use is restricted to a particular geographic location and needs a good amount of preprocessing for real-world data. After inversion, comes the task of picking faults and deciding the best possible spots to drill a well. The manual process of fault detection is time-consuming, labor-intensive, and subjective to experts' opinions. Automating this task can save a lot of time and facilitate better decision-making.

3. Automated Fault Interpretation

Fault detection and interpretation from the seismic images can be done by using computer vision. The state-of-the-art CNN can be put to use for this task. Real seismic images, along with augmented ones or synthetic data can be used to train the CNN model. There are different approaches to the problem [29]. One way could be to consider it as an Image Segmentation problem, in which each pixel can be classified as 0 (for non-fault) and 1(for fault). The other way is to consider it as an Image Classification problem and identify fault or non-fault just for the center of the image. The paper uses the later approach and also determines the fault dip and azimuth.

b. Data Analysis and Reservoir Modeling

i. Petrophysics

To maximize production, optimal production techniques are a must. Hence, understanding the behavior of reservoir fluids under different conditions is necessary. An important part of reservoir characterization is to construct 3D images of the petrophysical properties. Conventionally, this is done by using core, well logging and pressure test analyses. The results are physics-based and empirical models are used. Some relations are highly non-linear and simple empirical models might fail in giving better results. Also, these processes might take a great deal of time without assuring reliable results.

DL, on the other hand, can capture complex correlations and patterns among various features. Also, a lot of data is available in the form of core, well logs, seismic data and production data of development wells, which can be deployed in DL models. Determining petrophysical properties like facies, lithology, porosity, permeability, saturation, mobility, etc. can then become easier. Petrophysical properties can be estimated from pre-stack seismic data with the help of CNNs [30]. Two petrophysical properties, namely porosity and volume of clay are predicted. To train the CNN, two approaches are discussed. First, the end-to-end CNN that directly estimates the petrophysical properties from the pre-stack seismic data. Second, Cascaded CNNs that predict elastic properties at first, and then petrophysical properties are estimated using the elastic properties. Water saturation in carbonate reservoirs can also be determined using Artificial Neural Networks (ANNs) and ANFIS [31]. The input is conventional wireline log data, and the output is core Dean-Stark data. ANN can also be used for lithology estimation from spectral IP data. It is considered to be a multiclass classification problem. The input is spectral IP data. The output is four lithology labels: Intrusive rocks, Carbonate rocks, Skarn rocks, Mineralized rocks. Conventionally, this is done using equivalent circuit analysis [32]. However, incorrect circuit parameter selection can lead to unrealistic or divergent outcomes. Hence, neural networks provide a great way to deduce reasonable results.

A unique integrated DL solution has been presented for predicting petrophysics, pore pressure and geomechanics properties [33]. Three DL neural networks were trained. The first one predicted petrophysical properties. Compressional velocity, gamma ray, density, resistivity, and neutron logs were used as inputs to predict volumes of shale, sand, dolomite, calcite, kerogen, and porosity. The second neural network predicted pore pressure. The inputs in this example were compressional and shear velocity, density, resistivity, neutron logs, and porosity, shale volume, and kerogen volume predictions from the previous stage. The third neural network used only compressional velocity, shear velocity and density logs as input. Other petrophysical properties like permeability, fluid volume, kerogen properties, etc. have also been predicted using DL techniques.

For a large reservoir with enough log data, a deep neural network can be trained and the petrophysical characteristics for a completely new well can be estimated [34]. Thus, virtual logs at any location in the reservoir can be created by using information about the nearby wells. This is done with the help of Recurrent Neural Networks (RNN) and CNN.

ii. Reservoir Engineering

One of the most important tasks regarding a reservoir is reservoir calculations, to forecast the production and estimate the reserves. The forecasts assist in operational decision-making for the short term and help in the management of the reservoir in the long term. The traditional method involves fitting a curve through the history of the production volumes of a well, and extrapolating it to predict the future production, commonly called Decline Curve Analysis (DCA). A time-series forecasting approach using DL models can be used to do the same more efficiently.

1. Production Forecasting and Reservoir Estimation

RNNs can predict the production of a well using its history of the production and the history of production of the nearby wells as well [35]. The input features used were daily production data of oil, gas, and water, together with wellhead pressure. The output was multiphase production predictions in the near future. Different methods including DCA, physical models like IPR + VLP, DL models, and hybrid models (combining physical methods and DL), are compared to make short-term, mid-term, and long-term predictions [36]. DL models performed better than the rest in the short term whereas physical methods and hybrid models could forecast the long-term production better. These forecasts can be improved by using EnKF (Ensemble Kalman Filter) - enhanced RNNs [37]. Estimated Ultimate Recovery (EUR) can be estimated simply with the help of geological data [38]. It uses inputs as geological parameters like Thickness, Porosity, Bulk Density, Vitrinite Reflectance, Water Saturation, Total Organic Carbon, and Brittleness. The method used is called stacked denoising autoencoders.

2. Well Placement

Well Placement optimization is an important aspect of Reservoir Engineering. The goal is to maximize the economic benefits by extracting maximum hydrocarbon production. Conventionally, full-physics reservoir simulation (RS) has been used for estimating productivity. But, it is computationally expensive. DL techniques prove to be faster and more accurate than the reservoir simulators. CNNs with robust optimization can be used for determining the placement of an oil well at a reservoir [39]. The CNN takes near-wellbore permeability data as input. The output is cumulative oil production at a feasible well location. This output is maximized to obtain the best possible location for well placement. Thus, the CNN is trained to correlate the petrophysical spatial data with the productivity levels, which are used in well-placement optimization.

c. Production

1. Production Surveillance and Analysis

A lot of data is now available in the oil and gas industry because of the well sensors that record various information after every few seconds. Permanent downhole gauges (PDGs) continuously record information like pressure, temperature, and flow rates. Analysis of the PDG data plays an important role in production surveillance. One of the critical tasks is to develop a relationship between pressure and flow rate. Recurrent Neural Networks (RNNs) can be used to learn patterns from the PDG data [40]. Two models are implemented to study the flow rate and pressure relationships: Standard RNN model and Nonlinear autoregressive exogenous model (NARX). Both the models were efficient and studying and predicting the flowrate and pressure data. It also shows how flow rate can be modeled on the basis of temperature. The NARX model was successful in learning the mappings between temperature and the flow rate. This can be of immense importance from an engineering point of view as flow rate profiles can be constructed by using the temperature data from sensors. The same analysis can be done on the production history of the existing wells [41]. The production dataset used has information about average downhole pressure, average downhole temperature, oil flow rate, gas flow rate, water flow rate, on-stream hours and choke size percentage. This data is used for estimating the pressure by using methods like Gated Recurrent Unit (GRU), Long Short Term Memory (LSTM) and a combination of RNN with CNNs and LSTNet (Long Short Term Time-series Networks). The study helps in understanding the short-term and long-term recurring patterns.

2. Predictive Maintenance

The oil and gas sector relies on a variety of massive, expensive machines. Huge amounts of losses can occur if these aren't functional and well maintained. By employing predictive maintenance, these risks can be minimized. It includes monitoring the machines, predicting failures, maintenance decision-making and hence, cutting the operating costs caused by severe equipment failure. ANNs can be used to predict Remaining Useful Life (RUL) of the machinery used in the oil and gas industry, helping in better decision making related to maintenance costs and ensuring reliability of the machines [42]. A multivariate dataset for a duration of 1 year is used for training. An ANN model is then trained for RUL prediction, after filtering the important ones using correlation. Predictive maintenance for the air booster compressor motor, which is an important device in oil and gas processes can also be devised [43]. A prediction model that uses RNN-LSTM (Recurrent Neural Networks - Long Short Term Memory) has been trained and simulated to provide early warning on motor faults. Various parameters related to the motor like the current, active power, discharge temperature, winding and bearing temperature, airflow, air pressure and vibration are considered for prediction. DL algorithms can be employed for health state prognostics of physical assets in two real-life scenarios from the OGI [44]. The first scenario involves an offshore natural gas treatment plant where anomalous CO_2 levels in the treated gas are to be predicted. The input, received from the sensors located at various points in the plant, provide data related to non-treated gas flow rate, its temperature, pressure of the amine at the reboiler, etc. LSTM-based encoder is used for anomaly detection by forecasting the elevated levels of CO_2. In the second situation, the health state of a sea water injection centrifugal pump must be predicted. Sensor measurements provide information about temperature, flow, suction pressure, discharge pressure, etc. A CNN-LSTM model is trained for a multi-class classification and the health state (Normal, Incipient, Degraded, Critical) is predicted.

3. Real-time Hydraulic Fracture Analysis

Hydraulic fracturing is a well stimulation process in which large volumes of frac fluids are injected into the well under high pressure. This creates fractures in reservoirs and creates new pathways for extracting oil. For a successful hydraulic fracturing, it is important to monitor the subsurface fractures and the formation of these subsurface fractures is indicated by surface pressure profiles. Hence, ensuring that the surface pressure data is accurate and devoid of noise is extremely crucial. RNNs can be used to estimate the real-time surface pressure during hydraulic fracturing [45]. The surface pumping data, which includes surface pressure, pumping flow rate and proppant concentration, is used as input. The LSTM is used for predicting the surface pressure in real-time. Sometimes during the process of hydraulic fracturing, the solid proppant used in frac fluids forms a bridge across the perforations. As a result, fluid flow is suddenly and significantly restricted, leading to a quick rise in pump pressure, causing screen-out. DL, along with physics-based approaches, can provide advanced warning of screen-outs in real-time [46]. The input included 3 real-time time series and 11 engineering features. The real-time time series input had information about the surface treatment pressure, flow rate and proppant concentration. The engineering features covered cumulative flow rate, cumulative proppant volume, pressure divided by flow rate, geological areas and formation types. The models tested were CNN-LSTM and ensemble model.

2. Midstream

a. Pipeline Leak Detection

Pipeline leaks are prevalent in remote oil and gas sites. Because of the harsh environments in which oil and gas pipelines are placed, it is difficult to rely on human operators to physically monitor the pipelines and respond to leaks. With the changes in climate and the soil composition there is incessant wear and tear in the pipelines that hinders the optimal flow and reduces the efficiency of transportation [47]. The delay in detection of these wear and tear may result in leakage that can have a major impact on the environment, human life, and economy. Therefore, early detection and prevention of wear and tear, leakage and localization of faults is a challenging problem [48]. CNNs can be used to detect leakages in pipelines [49]. IoT cameras installed at various points on the pipeline capture images that are used as input for the CNNs. It involved two tasks, image classification (whether or not a leakage is present) and image localization (getting an exact location of the leak).

b. Pipeline Maintenance

Pigging is a technique used by operators to run oil and gas pipelines more smoothly. PIGs, which stands for Pipeline Inspection Gauge, are used to undertake various maintenance tasks. They are utilized for a variety of tasks, including pipeline cleaning and inspection. Analyzing pigging data is crucial, and it necessitates knowledge on how to utilize specialized software as well as the capacity to comprehend the signals gathered by the sensors. CNNs can be used on the images from the in-line inspection magnetic flux leakage data. The CNNs are trained on labeled data for Feature Detection and Metal Loss Detection [50].

3. Downstream

The downstream sector includes refining and marketing. One of the most difficult tasks in the oil refining and petrochemical sectors is maintaining production equipment while continuing to operate. As a result, technology that can foresee equipment breakdown is in high demand. After achieving a stable operation, the next goal is to increase production and maximize profitability. Because separation and other operations in the oil refining and petrochemical sectors consume a significant amount of energy, it is also critical to reduce energy consumption [51].

a. Predictive Analysis

Predictive analysis can help detect changes in system behavior far ahead of standard operational alarms, giving more time to correct problems. Various algorithms are being used for the process. Neural Networks use parametric technology but take a lot of time in the training process. Clustering identifies groups/clusters of data, for example, turbine cycles from start-up to shut down. Decision Tree Learning, Fuzzy Logic, DL and other Pattern Recognition techniques are also being used [52].

b. Quality Estimation

Evolving Intelligent sensors can be used for estimating various quality parameters of the distillation process. Some of the examples are discussed. The temperature of the heavy naphtha when it evaporates 95% liquid volume can be predicted by using inputs like pressure of the tower, amount of the product taking off, density of the crude, temperature of the column overhead and the temperature of the naphtha extraction. To predict the temperature of the kerosene when it evaporates 95% liquid volume, inputs like pressure of the tower, amount of product taking off, density of the crude, temperature of the column overhead, steam introduced in GOL stripper, temperature of the kerosene extraction and temperature of the naphtha extraction are used. The Abel Inflammability Index of the Kerosene can also be predicted using a set of relevant features [53].

CONCLUSION

The oil and gas industry has numerous hurdles to overcome in order to remain viable in the future. Many of its procedures have been assisted by the development of data-driven technologies. Although, various limitations can be identified in these applications. The results produced by the algorithms heavily rely on the inputs, which are unique to the region, machinery or the process involved. As a result, there is no universal algorithm structure that can be utilized in all circumstances; rather, the algorithm must be modified each time. DL approaches work effectively with Big Data. The data obtained for some of the tasks related to OGI is limited. Hence, deploying deep learning algorithms on such small datasets makes it difficult to acquire useful insights. Furthermore, DL algorithms such as CNN and LSTM were developed with a specific objective. These models' specifications and structure are unique to the task at hand. For better results, a new application necessitates a model that is essentially being designed for it. Additionally, the data obtained from the industry is not publicly accessible. To overcome these limitations, several steps can be taken. First, developing models that cater to the needs of a particular task in the oil and gas industry that needs to be solved, which would require expertise in both domain knowledge and deep learning. Using Transfer Learning and making data accessible so that more researchers can work upon it can also be helpful. Also, methods can be developed to acquire more data. The industry faces newer challenges while embracing ML and DL. The gap between data scientists and petroleum engineers needs to be bridged. Customizing the deep learning techniques according to the need of the problem creates a demand for individuals who have both, data science skills as well as domain expertise. To implement machine learning and deep learning techniques, it is important to monitor the facilities continuously. There is a need for continuous information related to the reservoir, wells and facilities. But the resources to collect this information as well as software and hardware requirements for doing so are limited. The industry currently lacks in this aspect. This necessitates the need of advanced technologies, sensors and softwares that can provide and process a stream of data in an efficient and accurate manner. Furthermore, the various processes in upstream like exploration, production, and reservoir management are quite disconnected. The absence of synchronicity between these activities makes it difficult to develop global strategies to improve the industry. Data from one process may be useful in anticipating the results of another process, hence there should be a global platform that integrates data from all processes. Real-time insights drawn simultaneously from hundreds of wells can be used for decision-making regarding the maintenance and operations of the reservoir. There are a few barriers like high research and development costs, aversion to change and high levels of risk. With open collaboration and knowledge sharing mindset at an industry-level, new opportunities can be identified easily, and existing facilities can be managed effectively. Collaboration starts with integrating the domains of the industry, like promoting collaborative efforts from geologists, geophysicists, petrophysicists, reservoir engineers. Next step is to collaborate on the planning and operations side. Once the processes and systems of a firm are integrated, the next step is to promote collaboration among the competitors in the industry. Open source is the key to innovation in the software industry. If the data is made available to all the stakeholders and the academia, the pace of innovation will definitely increase. If collaborative efforts are shown by the competitors in the industry, then the growth of AI-assisted decision making is certain.

REFERENCES

[1] H. Lu, K. Huang, M. Azimi, and L. Guo, "Blockchain Technology in the Oil and Gas Industry: A Review of Applications, Opportunities, Challenges, and Risks," *IEEE Access*, vol. 7, pp. 41426–41444, 2019, doi: 10.1109/ACCESS.2019.2907695.

[2] B. Lin and R. Bai, "Oil prices and economic policy uncertainty: Evidence from global, oil importers, and exporters' perspective," *Res. Int. Bus. Financ.*, vol. 56, p. 101357, Apr. 2021, doi: 10.1016/j.ribaf.2020.101357.

[3] M. I. Haque, "Oil price shocks and energy consumption in GCC countries: a system-GMM approach," *Environ. Dev. Sustain.*, vol. 23, no. 6, pp. 9336–9351, Jun. 2021, doi: 10.1007/s10668-020-01027-y.

[4] H. Alqaralleh, "On the asymmetric response of the exchange rate to shocks in the crude oil market," *Int. J. Energy Sect. Manag.*, vol. 14, no. 4, pp. 839–852, Mar. 2020, doi: 10.1108/IJESM-10-2019-0011.

[5] R. J. Kilian, Lutz and Vigfusson, "The Role of Oil Price Shocks in Causing U.S. Recessions," *J. Money, Credit Bank.*, vol. 49, no. 8, pp. 1747–1776, Dec. 2017, doi: 10.1111/jmcb.12430.

[6] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, "Application of machine learning and artificial intelligence in oil and gas industry," *Pet. Res.*, Jun. 2021, doi: 10.1016/j.ptlrs.2021.05.009.

[7] D. Al-Subaiei, M. Al-Hamer, A. Al-Zaidan, and M. S. Nawaz, "Smart Production Surveillance: Production Monitoring and Optimization Using Integrated Digital Oil Field," Oct. 2019, doi: 10.2118/198114-MS.

[8] I. S. Korovin, S. Sisyukin, and G. Schaefer, "Application of Convolutional Neural Networks for Dynamic Digital Oil Field Models," 2019, pp. 1099–1106.

[9] F. A. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization," *J. Pet. Sci. Eng.*, vol. 151, pp. 480–487, Mar. 2017, doi: 10.1016/j.petrol.2017.01.024.

[10] H. Lu, L. Guo, M. Azimi, and K. Huang, "Oil and Gas 4.0 era: A systematic review and outlook," *Comput. Ind.*, vol. 111, pp. 68–90, Oct. 2019, doi: 10.1016/j.compind.2019.06.007.

[11] K. M. Hanga and Y. Kovalchuk, "Machine learning and multi-agent systems in oil and gas industry applications: A survey," *Comput. Sci. Rev.*, vol. 34, p. 100191, Nov. 2019, doi: 10.1016/j.cosrev.2019.08.002.

[12] R. N. Anderson, "'Petroleum Analytics Learning Machine' for optimizing the Internet of Things of today's digital oil field-to-refinery petroleum system," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec. 2017, pp. 4542–4545, doi: 10.1109/BigData.2017.8258496.

[13] F. Anifowose, A. Abdulraheem, and A. Al-Shuhail, "A parametric study of machine learning techniques in petroleum reservoir permeability prediction by integrating seismic attributes and wireline data," May 2019. doi: 10.1016/j.petrol.2019.01.110.

[14] J. Zheng *et al.*, "Deeppipe: A semi-supervised learning for operating condition recognition of multi-product pipelines," *Process Saf. Environ. Prot.*, vol. 150, pp. 510–521, Jun. 2021, doi: 10.1016/j.psep.2021.04.031.

[15] N. Brown, A. Roubíčková, I. Lampaki, L. MacGregor, M. Ellis, and P. Vera de Newton, "Machine learning on Crays to optimize petrophysical workflows in oil and gas exploration," *Concurr. Comput. Pract. Exp.*, vol. 32, no. 20, Oct. 2020, doi: 10.1002/cpe.5655.

[16] M. Mohammed, M. B. Khan, and E. B. M. Bashier, *Machine learning: algorithms and applications*. Crc Press, 2016.

[17] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, Feb. 2020, doi: 10.1007/s10994-019-05855-6.

[18] D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," *J. Pet. Sci. Eng.*, vol. 200, p. 108182, May 2021, doi: 10.1016/j.petrol.2020.108182.

[19] M. Ippolito, J. Ferguson, and F. Jenson, "Improving facies prediction by combining supervised and unsupervised learning methods," *J. Pet. Sci. Eng.*, vol. 200, p. 108300, May 2021, doi: 10.1016/j.petrol.2020.108300.

[20] K. Zhang, N. Lin, C. Fu, D. Zhang, X. Jin, and C. Zhang, "Reservoir characterisation method with multi-component seismic data by unsupervised learning and colour feature blending," *Explor. Geophys.*, vol. 50, no. 3, pp. 269–280, May 2019, doi: 10.1080/08123985.2019.1603078.

[21] M. Compare, L. Bellani, E. Cobelli, and E. Zio, "Reinforcement learning-based flow management of gas turbine parts under stochastic failures," *Int. J. Adv. Manuf. Technol.*, vol. 99, no. 9–12, pp. 2981–2992, Dec. 2018, doi: 10.1007/s00170-018-2690-6.

[22] F. Hourfar, H. J. Bidgoly, B. Moshiri, K. Salahshoor, and A. Elkamel, "A reinforcement learning approach for waterflooding optimization in petroleum reservoirs," *Eng. Appl. Artif. Intell.*, vol. 77, pp. 98–116, Jan. 2019, doi: 10.1016/j.engappai.2018.09.019.

[23] L. Zhu *et al.*, "Forming a new small sample deep learning model to predict total organic carbon content by combining unsupervised learning with semisupervised learning," *Appl. Soft Comput.*, vol. 83, p. 105596, Oct. 2019, doi: 10.1016/j.asoc.2019.105596.

[24] E. Elyan, L. Jamieson, and A. Ali-Gombe, "Deep learning for symbols detection and classification in engineering drawings," *Neural Networks*, vol. 129, pp. 91–102, Sep. 2020, doi: 10.1016/j.neunet.2020.05.025.

[25] J. D. Salazar *et al.*, "3D PHOTOGRAMMETRIC INSPECTION OF RISERS USING RPAS AND DEEP LEARNING IN OIL AND GAS OFFSHORE PLATFORMS," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLIII-B2-2, pp. 1265–1272, Aug. 2020, doi: 10.5194/isprs-archives-XLIII-B2-2020-1265-2020.

[26] W. Chen, L. Yang, B. Zha, M. Zhang, and Y. Chen, "Deep learning reservoir porosity prediction based on multilayer long short-term memory network," *GEOPHYSICS*, vol. 85, no. 4, pp. WA213–WA225, Jul. 2020, doi: 10.1190/geo2019-0261.1.

[27] Y. N. Pandey, A. Rastogi, S. Kainkaryam, S. Bhattacharya, and L. Saputelli, *Machine Learning in the Oil and Gas Industry*. 2020.

[28] D. A. B. Oliveira, R. S. Ferreira, R. Silva, and E. Vital Brazil, "Interpolating Seismic Data with Conditional Generative Adversarial Networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1952–1956, 2018, doi: 10.1109/LGRS.2018.2866199.

[29] Y. Zheng, Q. Zhang, A. Yusifov, and Y. Shi, "Applications of supervised deep learning for seismic interpretation and inversion," *Lead. Edge*, vol. 38, no. 7, pp. 526–533, 2019, doi: 10.1190/tle38070526.1.

[30] V. Das and T. Mukerji, "Petrophysical properties prediction from prestack seismic data using convolutional neural networks," *Geophysics*, vol. 85, no. 5, pp. N41–N55, 2020, doi: 10.1190/geo2019-0650.1.

[31] M. Rasheed Khan, Z. Tariq, and A. Abdulraheem, "Machine learning derived correlation to determine water saturation in complex lithologies," *Soc. Pet. Eng. - SPE Kingdom Saudi Arab. Annu. Tech. Symp. Exhib. 2018, SATS 2018*, 2018, doi: 10.2118/192307-ms.

[32] S. Shin, Y. Li, S. Park, and S. J. Cho, "Supervised machine learning for lithology estimation using spectral induced polarization data," *2018 SEG Int. Expo. Annu. Meet. SEG 2018*, pp. 1898–1902, 2018, doi: 10.1190/segam2018-2998355.1.

[33] E. Z. Naeini, S. Green, I. Russell-Hughes, and M. Rauch-Davies, "An integrated deep learning solution for petrophysics, pore pressure, and geomechanics property prediction," *Lead. Edge*, vol. 38, no. 1, pp. 53–59, 2019, doi: 10.1190/tle38010053.1.

[34] M. Korjani, A. Popa, E. Grijalva, S. Cassidy, and I. Ershaghi, "A new approach to reservoir characterization using deep learning neural networks," *Soc. Pet. Eng. - SPE West. Reg. Meet.*, 2016, doi: 10.2118/180359-ms.

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

[36] P. S. Omrani *et al.*, "Deep learning and hybrid approaches applied to production forecasting," *Soc. Pet. Eng. - Abu Dhabi Int. Pet. Exhib. Conf. 2019, ADIP 2019*, 2019, doi: 10.2118/197498-ms.

[37] A. Bao, E. Gildin, J. Huang, and E. J. R. Coutinho, "Data-driven end-to-end production prediction of oil reservoirs by EnKF-enhanced recurrent neural networks," *SPE Lat. Am. Caribb. Pet. Eng. Conf. Proc.*, vol. 2020-July, 2020, doi: 10.2118/199005-ms.

[38] L. Crnkovic-Friis and M. Erlandson, "Geology driven EUR prediction using deep learning," *Proc. - SPE Annu. Tech. Conf. Exhib.*, vol. 2015-Janua, no. September, pp. 1062–1071, 2015, doi: 10.2118/174799-ms.

[39] S. Kwon, G. Park, Y. Jang, J. Cho, M. gon Chu, and B. Min, "Determination of oil well placement using convolutional neural network coupled with robust optimization under geological uncertainty," *J. Pet. Sci. Eng.*, vol. 201, p. 108118, 2021, doi: 10.1016/j.petrol.2020.108118.

[40] C. Tian and R. N. Horne, "Recurrent neural networks for permanent downhole gauge data analysis," *Proc. - SPE Annu. Tech. Conf. Exhib.*, vol. 0, no. October, pp. 9–11, 2017, doi: 10.2118/187181-ms.

[41] Y. Li, R. Sun, and R. Horne, "Deep learning for well data history analysis," *Proc. - SPE Annu. Tech. Conf. Exhib.*, vol. 2019-Septe, no. October, 2019, doi: 10.2118/196011-ms.

[42] M. F. A. M. Fauzi, I. A. Aziz, and A. Amiruddin, "The Prediction of Remaining Useful Life (RUL) in Oil and Gas Industry using Artificial Neural Network (ANN) Algorithm," *2019 IEEE Conf. Big Data Anal. ICBDA 2019*, pp. 7–11, 2019, doi: 10.1109/ICBDA47563.2019.8987015.

[43] T. Abbasi, K. H. Lim, and K. S. Yam, "Predictive Maintenance of Oil and Gas Equipment using Recurrent Neural Network," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 495, no. 1, 2019, doi: 10.1088/1757-899X/495/1/012067.

[44] J. Figueroa Barraza, L. Guarda Bräuning, R. Benites Perez, C. B. Morais, M. R. Martins, and E. L. Droguett, "Deep learning health state prognostics of physical assets in the Oil and Gas industry," *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, 2020, doi: 10.1177/1748006X20976817.

[45] S. Madasu and K. P. Rangarajan, "Deep recurrent neural network DRNN model for real-time multistage pumping data," *OTC Arct. Technol. Conf. 2018*, 2018, doi: 10.4043/29145-ms.

[46] J. J. Sun, A. Battula, B. Hruby, and P. Hossaini, "Application of Both Physics-Based and Data-Driven Techniques for Real-Time Screen-Out Prediction with High Frequency Data," no. Schlumberger, 2020, doi: 10.15530/urtec-2020-3349.

[47] S. Choubey and G. P. Karmakar, "Artificial intelligence techniques and their application in oil and gas industry," *Artif. Intell. Rev.*, vol. 54, no. 5, pp. 3665–3683, Jun. 2021, doi: 10.1007/s10462-020-09935-1.

[48] F. Idachaba and M. Rabiei, "Current Technologies and the Applications of Data Analytics For Crude Oil Leak Detection in Surface Pipelines," *J. Pipeline Sci. Eng.*, Nov. 2021, doi: 10.1016/j.jpse.2021.10.001.

[49] O. Ibitoye, O. Shafiq, and A. Matrawy, "A Convolutional Neural Network Based Solution for Pipeline Leak Detection Poster Abstract: A Convolutional Neural Network Based Solution for Pipeline Leak Detection," no. October, pp. 4–7, 2019, [Online]. Available: https://www.researchgate.net/publication/337060339.

[50] V. Nidhi, R. Rao, and A. Mittal, "AI based intelligent pigging solution," *Soc. Pet. Eng. - Abu Dhabi Int. Pet. Exhib. Conf. 2019, ADIP 2019*, no. November, pp. 11–14, 2019, doi: 10.2118/197723-ms.

[51] T. Ohtani, "Application of AI to Oil Refineries and Petrochemical Plants," *Yokogawa Tech. Rep. English Ed.*, vol. 63, no. 1, pp. 7–10, 2020.

[52] M. A. Turk and D. Ph, "Predictive Analytics to Monitor the Health of Refinery Rotating Equipment."

[53] J. Macías-Hernández and P. Angelov, "Applications of Evolving Intelligent Systems to Oil and Gas Industry," in *Evolving Intelligent Systems: Methodology and Applications*, 2010, pp. 401–421.