

## TOPICAL REVIEW

# Machine Learning in Oil and Gas Exploration: A Review

**AHMAD LAWAL<sup>1</sup>**, **YINGJIE YANG<sup>1</sup>**, **HONGMEI HE<sup>2</sup>**, AND **NATHANAEL L. BAISA<sup>1</sup>**<sup>1</sup>School of Computer Science and Informatics, De Montfort University, LE1 9BH Leicester, U.K.<sup>2</sup>School of Science, Engineering and Environment, University of Salford, M5 4WT Manchester, U.K.

Corresponding author: Ahmad Lawal (ahmad.lawal@dmu.ac.uk)

**ABSTRACT** A comprehensive assessment of machine learning applications is conducted to identify the developing trends for Artificial Intelligence (AI) applications in the oil and gas sector, specifically focusing on geological and geophysical exploration and reservoir characterization. Critical areas, such as seismic data processing, facies and lithofacies classification, and the prediction of essential petrophysical properties (e.g., porosity, permeability, and water saturation), are explored. Despite the vital role of these properties in resource assessment, accurate prediction remains challenging. This paper offers a detailed overview of machine learning's involvement in seismic data processing, facies classification, and reservoir property prediction. It highlights its potential to address various oil and gas exploration challenges, including predictive modelling, classification, and clustering tasks. Furthermore, the review identifies unique barriers hindering the widespread application of machine learning in the exploration, including uncertainties in subsurface parameters, scale discrepancies, and handling temporal and spatial data complexity. It proposes potential solutions, identifies practices contributing to achieving optimal accuracy, and outlines future research directions, providing a nuanced understanding of the field's dynamics. Adopting machine learning and robust data management methods is crucial for enhancing operational efficiency in an era marked by extensive data generation. While acknowledging the inherent limitations of these approaches, they surpass the constraints of traditional empirical and analytical methods, establishing themselves as versatile tools for addressing industrial challenges. This comprehensive review serves as an invaluable resource for researchers venturing into less-charted territories in this evolving field, offering valuable insights and guidance for future research.

**INDEX TERMS** Oil and gas exploration, machine learning, petrophysical properties prediction, facies and lithofacies classification, seismic data processing.

## I. INTRODUCTION

The oil and gas industry is a sophisticated sector that combines many complex activities in its value chain broadly segmented into Upstream, Midstream, and Downstream, as illustrated in Fig. 1. In any industry operation, an unprecedented amount of data can be generated from the equipment involved and routine human logs. The Upstream segment, which concerns the exploration and production of oil and natural gas, produces data such as geological surveys, well logs, and readings from drilling equipment. This segment

is also expected to generate significantly higher volumes of data with improvements in seismic acquisition devices, channel counting, and fluid front monitoring geophones [1]. The midstream segment involves transporting and storing crude oil and natural gas using pipelines and their associated infrastructure such as pumping stations and, tank trucks, etc. All these enable the generation of large volumes of data. The downstream segment involves turning crude oil and natural gas into finished products and marketing them accordingly. This involves generating and analyzing large amounts of data for competitive advantage and cost reduction. The amount of data generated in the oil and gas industry is so enormous that even its capture and storage requires sophisticated techniques

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos<sup>1</sup>.

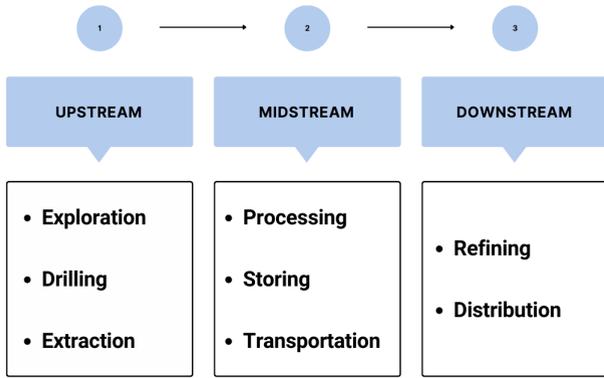


FIGURE 1. Oil and gas production process.

and expertise, let alone analysis, to derive hidden actionable insights.

Oil and Gas Exploration is the practice of attempting to locate accumulations of oil and natural gas trapped under the surface of the Earth’s atmosphere by utilizing petroleum geology. Exploration is carried out to offer the knowledge necessary to use the best prospects presented by the regions chosen for exploration and to oversee the research operations on the blocks that have been obtained. Exploration controls the inherent risks involved in this process and is generally handled by selecting various probabilistic and economically favourable options.

Procedures are commonly used in oil and gas exploration to locate, evaluate, and exploit hydrocarbon resources. Identifying and acquiring promising locations is the initial phase, which may include studying geological information and conducting aerial surveys to identify regions with a high likelihood of harbouring hydrocarbon resources. A geological survey should be undertaken better to understand the geology and hydrocarbon potential of the area when a promising location is identified. This often uses various methods, such as electromagnetic, magnetic, seismic, and gravity surveys. Seismic surveys are one of the most essential methods used to search for oil and gas. They work by sending sound waves to the ground and recording and analyzing their reflections. The collected data may provide extensive information on underlying geology and aid in identifying the probable hydrocarbon sources.

After a possible reservoir has been located, exploratory drilling may be carried out to assess the reservoir’s existence, quality, and amount of hydrocarbons. Drilling one or more exploratory wells to collect core samples, fluid samples, and other data that may be studied to identify the reservoir’s properties is routine.

If hydrocarbons were identified, the next stage was to assess the project’s economic feasibility. This includes determining the reservoir size and productivity and the costs of drilling, production, and transportation. If the idea is deemed commercially feasible, the field will be developed, and production will commence. The construction

of production facilities, drilling of production wells, and use of various technologies and procedures to improve output and maximum recovery are expected.

Overall, the stages involved in oil and gas exploration are complicated and require various technical skills and resources. A successful exploration operation, on the other hand, may lead to the finding of significant hydrocarbon deposits that can supply substantial energy sources for humanity. The stages are shown in Fig. 2.



FIGURE 2. Stages in oil and gas exploration.

Although risks cannot be eliminated, they may be managed and reduced using appropriate operational, conceptual, and technological breakthroughs such as reservoir characterization. Reservoir characterization quantitatively defines different reservoir features regarding their geographic variability by integrating data collected from the field and laboratory. It is a crucial aspect of the management of emerging reservoirs. Reservoir characterization provides more insight into the reservoir and its behaviour, which helps detect possible drilling risks and improves the ability to recommend well placement.

Employing a data-driven approach to address problems in the development process of oil and gas exploration and production is not a new concept, as it surpasses the limitations posed by traditional techniques. Machine learning has been used to address problems such as regression, classification, and function approximations. Traditional methods are typically redundant and time-consuming and rely on trial and error to achieve optimum results. They cannot accommodate missing data or background noise and fail to perform efficiently when presented with overwhelming interdependencies, requiring several simplifications and biased assumptions. Data-driven procedures were utilized to overcome these problems. Data-driven approaches provide methodologies that incorporate various data formats, calculate uncertainty, discover hidden patterns, and extract the relevant data. This data type is critical for estimating future trends, resolving challenges, and anticipating unexpected activities using traditional procedures. Data-driven predictions and decisions are made using machine learning that accepts extensive data. Machine learning has been used to address problems, including regression, classification and function approximation, in the development process for oil and gas exploration and production.

Significant progress has been made in this area, and there have been a number of reviews. Existing reviews in the field of machine learning in the oil and gas industry tend to offer a broad, high-level perspective [2], [3], [4], [5], there is room for further exploration to delve into the intricate challenges and nuances specific to the exploration stage in this complex sector. Some of these challenges encompass the inherent

uncertainties in various subsurface exploration parameters, scale discrepancies and the complexities related to handling temporal and spatial data in exploration processes. This limited scope results in a gap in addressing the specific hurdles encountered across various industry sectors. Furthermore, while some reviews touch upon the challenges inherent in applying machine learning in the broader oil and gas domain, they frequently fail to provide potential solutions or guide future research endeavours.

This review aims to provide a comprehensive and up-to-date overview of machine learning applications in upstream oil and gas exploration. It aims to highlight the potential of machine learning to address various challenges in this field, identify key barriers impeding its widespread application, and offer potential development trends and identify practices that contribute to achieving optimal accuracy. The review also outlines future research directions, providing a nuanced understanding of the field's dynamics.

This review brings novelty through three key dimensions. Firstly, it delivers a deeply comprehensive study of machine learning in the industry's exploration phase, specifically focusing on geological and geophysical aspects. When it comes to exploration, critical areas need to be addressed such as seismic data processing, lithofacies classification, and predicting petrophysical properties. These areas come with their own unique challenges, including inherent uncertainties in various subsurface exploration parameters, discrepancies in scale, and complexities related to handling temporal and spatial data. This review provides a holistic view of how machine learning is harnessed in the industry by encompassing a broad spectrum of topics.

Furthermore, the review adeptly identifies and discusses emerging trends in machine learning applications. It casts a spotlight on the latest developments and innovations within the field, shedding light on how these trends actively shape the future of upstream oil and gas exploration. This forward-looking approach ensures that the review captures the current state of the art and provides valuable insights into the industry's potential evolution.

Lastly, the review stands out for its pragmatic approach to addressing successes and challenges. While celebrating the accomplishments of machine learning in the oil and gas sector, it does not shy away from highlighting critical issues such as data issues, model interpretability, and deployment complexities. Furthermore, this comprehensive review provides potential solutions and recommended practices that contribute to achieving optimal accuracy to address these challenges effectively while highlighting promising avenues for future research. This balanced perspective equips readers with a nuanced understanding of the field's dynamics and the means to navigate them effectively.

This article explores the application of machine learning in addressing challenges within the upstream oil and gas industry, with a focus on exploration. Section II outlines the review's methodology. Section III delves into seismic data processing and lithofacies classification in geological

and geophysical exploration, while Section IV covers the prediction of petrophysical properties in reservoir characterization. In Section VII, we discuss the strengths and weaknesses of existing machine learning strategies for these issues, presenting a roadmap for optimal accuracy in their applications. Section VII outlines current challenges, proposes solutions, and identifies future research directions.

The highlights of this review are stated below:

- **Comprehensive Coverage:** The paper offers an extensive overview of machine learning applications within the exploration stage of upstream oil and gas.
- **Key Focus Areas:** It explores seismic data processing, lithofacies classification, and prediction of petrophysical properties such as porosity, permeability, and water saturation.
- **Identification of Barriers:** The paper identifies unique challenges and limitations that hinder the widespread adoption of machine learning in the exploration sector.
- **Potential Solutions:** It provides potential solutions and identifies practices that contribute to achieving optimal accuracy to address the identified challenges effectively.
- **Balanced Approach:** The paper takes a balanced approach by acknowledging the achievements of machine learning while addressing critical issues like data issues and model interpretability.
- **Guidance for Future Research:** It outlines future research directions, offering a roadmap for those interested in the industry's evolving landscape of machine learning.

## II. METHODOLOGY

The methodology presents a comprehensive overview of the approach employed for the literature review focused on machine learning applications within the upstream oil and gas sector. The methodology outlined here serves as the foundation for systematically identifying, selecting, and critically analyzing relevant studies in the field. We aim to give readers insights into our approach's robustness and rigour, ensuring the review's credibility and comprehensiveness.

### A. SELECTION OF RELEVANT LITERATURE

This stage details our search strategy, keywords, and databases used. It also outlines the criteria for selecting pertinent literature.

For this literature review, we meticulously followed a systematic approach to identify and include studies related to machine learning applications in upstream oil and gas, focusing on geological exploration and reservoir characterization. We conducted comprehensive searches across reputable academic databases, including IEEE Xplore, ScienceDirect, and Google Scholar. Our search queries incorporated carefully selected keywords, such as "lithofacies classification", "machine learning", "upstream oil and gas", "geological exploration", "permeability prediction", "porosity prediction", "water saturation prediction", "reservoir characterization", and "seismic data processing".

We established specific inclusion criteria to ensure the quality and relevance of the studies in our review. These criteria encompassed relevance to machine learning applications in geological exploration and reservoir characterization in the upstream oil and gas sector, adherence to rigorous peer-reviewed standards, and publication in English. As a result, we narrowed our selection to a total of 128 papers for the review.

### B. DATA COLLECTION AND SYNTHESIS

This stage details the data extraction process from the relevant literature and how the selected literature was categorized to enhance the structure of the findings.

In this methodology phase, we conducted a detailed analysis of the chosen literature. This analysis involved extracting essential details from each study, including research objectives, methodologies, key findings, and limitations. The goal was to create a comprehensive dataset from the literature, providing a well-rounded perspective for our review.

Subsequently, we systematically categorized the literature into coherent themes, such as seismic data processing, facies classification, and prediction of petrophysical properties. This thematic organization allowed us to present the collective findings in a structured manner, facilitating the identification of common trends and patterns across the literature.

### C. CRITICAL EVALUATION AND PRESENTATION

This stage critically assesses research quality, methodologies, and contributions, highlighting strengths, limitations, and findings organized by themes for clarity.

During the last stage of our methodology, we thoroughly evaluated each study's quality, research methodologies, and contributions to the field. We considered the machine learning techniques, data preprocessing strategies, feature selection methods, and model evaluation approaches. This critical analysis provides readers with insights into the strengths and limitations of the existing body of research. Our systematic thematic structure serves as a clear framework for presenting our findings, ensuring comprehensibility and providing valuable insights for our readers.

## III. GEOLOGICAL AND GEOPHYSICAL EXPLORATION

Geological and geophysical exploration is carried out using surface techniques to evaluate the physical characteristics of the underlying earth, coupled with variations in these qualities, to identify or deduce the existence and location of hydrocarbons (oil and gas) in economical amounts. This is done using physical methods, such as seismic, electrical, coring, and well logging methods, to evaluate the physical properties of rocks and, more specifically, to identify the measurable physical differences between rocks containing hydrocarbons and those that do not. This is helpful in the placement of offshore structures and in making knowledgeable decisions regarding the strategic and economic considerations of oil and gas operations.

### A. SEISMIC DATA PROCESSING

The primary geophysical approach employed to map geological features under the Earth's surface, whether on land or in marine environments, is seismic data. Inherently, human-driven interpretation processes are sluggish, costly, and non-reproducible. One of the most time-consuming activities is the interpretation of large amounts of seismic data. Because of their vast bulk, seismic data sets are well suited for sophisticated machine learning algorithms such as Convolutional Neural Networks (CNN), which must be sufficiently trained with substantial data to work efficiently and accurately. Several complex geological problems, including fault detection, salt-body identification, sweet spots, and seismic horizons, have been solved using machine learning with seismic data. Furthermore, although humans excel at discovering characteristics exclusively in two dimensions, well-written algorithms can function in all dimensions. The application of the Artificial Neural Network (ANN) technique in the field of exploration has produced fruitful outcomes in reducing exploration risks and increasing the efficiency of exploration wells [6].

Structural breaks may be caused by various types of subsurface movements, which can lead to the formation of faults. After considering the existence of defects in the area of interest, specific choices regarding operations must be made. In traditional processes, fault interpretation is a process that takes a significant amount of time. Henceforth, Guitton et al. [7] employed a Support Vector Machine (SVM) technique to detect faults in seismic sections. From the labelled seismic sections, the authors employed the Scale Invariant Feature Transform (SWIFT) and Histogram of Oriented Gradients (HOG) to extract a set of features that will be used to train the SVM to identify faults. The advantage of combining HOG and SIFT features has been noted as it surpasses their individual usage. However, Xiong et al. [8] revealed the weakness of SVM in demanding the precomputation of characteristics for mapping faults. A laborious process of manually mapping faults must be performed for every data set in the training dataset. In addition, the technique has poor performance in the zones with poor reflections. Hence, the superiority of CNN was presented in [9], [10], and [8]. Using 3D seismic data, Xiong et al. [8] CNN method automatically identified and mapped fault zones, eliminating the need for human precomputation.

Yang and Sun [11] suggested a technique for tracking horizons with complicated seismic reflection characteristics using deep CNNs. The suggested approach can determine the locations of faults and precisely extract horizons that cut across faults. The suggested model was more consistent and faster than the conventional 3D horizon tracking technique. The CNN-based approach has shown significant promise for enhancing the effectiveness and accuracy of horizon monitoring.

For a seismic full wave tomography study, Diersen et al. [12] suggested an ANN and Importance-Aided Neural

Network (IANN). The proposed models integrate machine learning and Complex Wavelet Transform (CWT), which is promising for improving the classification precision and speeding up the computation of the classification of observed data wave segments and synthesised data wave segment matches. Both ANN and IANN showed positive results, with IANN performing marginally better.

Using multiple neural network models, a considerable amount of 3D seismic data was processed by Rastegarnia et al. [13] to obtain electrofacies volumes and the 3D flow zone index (FZI). The authors suggested a probabilistic neural network (PNN) for the electrofacies model that uses multi-resolution graph-based clustering (MRGC) as an optimizer. The 3D FZI model, on the other hand, used a multi-attribute method utilising a radial basis function (RBF) network, a multilayer feed-forward network (MLFFN), and a PNN to enhance the model. According to the results, the two models are in excellent agreement with one another, and the PNN-based models can be used to estimate both the FZI and electrofacies volume efficiently.

In their study, La Marca et al. [14] introduced a novel quantitative assessment approach for unsupervised machine learning algorithms, employing techniques like Kmeans, Generative Topographic Maps (GTM), and Principal Component Analysis (PCA) in seismic interpretation. Their methodology, demonstrated using synthetic multi-dimensional seismic data, successfully clustered data into geologically meaningful groups. Machine learning expands the range of attributes analyzed and reveals intricate details often missed by human interpreters. It's noteworthy that machine learning algorithms are typically calibrated using well logs; nevertheless, human expertise also plays a pivotal role in the interpretation process.

Another study by Qian et al. [15] developed a support vector machine method by combining data from geology, well drilling, logging, and seismic surveys to make a multi-attribute estimate of reservoir sweet spots and conduct a thorough quantitative characterization of shale reservoirs. This technique was superior to conventional techniques by providing an efficient and accurate quantitative assessment approach for evaluating shale reservoirs.

## B. FACIES AND LITHOFACIES CLASSIFICATION

Facies are a basic geologic characteristic that influences hydrocarbon production, making rock facies understanding vital in oil and gas exploration [16]. Core and advanced well log data may provide this information, but access to this type of rich data is restricted by the expense and time required to acquire it. Numerous low-cost data-driven machine learning techniques leveraging inexpensive well log data have been proposed for subsurface research. These well log data advantages include continuous availability along depth and easy data collection. As a result, they constitute a valuable source of information about subsurface rock. Lithofacies are often determined by integrating petrophysical

and geological properties, and they can be an essential tool for reservoir characterisation [17]. Several mathematical approaches have been developed since the advent of well logs in predicting lithology relying on well logs [18]. Lithofacies classification is regularly done utilising core samples and wireline log data with machine learning.

Recently, machine learning has been utilised to assist in the labour-intensive evaluation of well logs for lithofacies classification. It can be used to classify lithofacies in uncored wells after being trained using other cored wells in the region. To train the model, lithofacies classification is applied to depth measurement based on the combination of well log and core data [19]. Gamma-ray (GR), resistivity(Rt), neutron(NPHI) density(RHOB), and lithology are the most often used logs for facies identification. These logs facilitate the generation of sophisticated characteristics that can improve predictions, including total organic matter(TOC) and matrix grain density(RHOMAA) [20].

Researchers have employed various machine learning techniques such as Neural Network(NN), SVM, and Random Forest(RF) [17], [19], [21], [22], [23], [24], [25], [26], [27]. They identified lithofacies from well logs across various reservoir types and, according to their findings, inferred that these strategies are effective and automated, requiring less time and resources than conventional methods.

Various Researchers [17], [19], [21], [22] carried out their research using NN to classify lithofacies. They concluded that it was excelling over the traditional methods. But when dealing with a small amount of data, Sebtosheikh and Salehi [23] noted that the SVM performs better. However, Xie et al. [16] conducted their research using both NN and SVM and deduced that both methods are affected by the number of features available, giving them a setback when limited features are used. The work established that ensemble methods are superior. The ensemble methods integrate numerous base models to create a single best prediction model [25]. Dell'aversana and, Tewari and Dwivedi [24], [26] also support the ensembles method as more robust, reliable and accurate. In another research, Hou et al. [28] compared Multilayer perception (MLP), SVM and ensemble eXtreme Gradient Boosting (XGboost) and RF models for lithofacies classification in the Gulong Shale. Based on the performance of the models, it can be concluded that the ensemble yields greater accuracy.

Despite this, new research has revealed that the Gradient Boosting (GB) approach outperforms other machine learning algorithms, mostly due to its robustness [19]. However, when working with large data sizes, RF outperforms. Bhattacharya and Mishra [27] studies similarly gave RF superiority over GB as it minimised the computing time during the training stage.

In a comparative study conducted by Al-Mudhafar et al. [29], various boosting algorithms were evaluated for lithofacies classification in an Iraqi carbonate reservoir. The study examined the performance of several boosting algorithms, including Logistic Boosting Regression (LogitBoost), Gen-

eralized Boosting Modelling (GBM), XGBoost, Adaptive Boosting Model (AdaBoost), and K-nearest neighbour (KNN), using input data derived from well logs and core data. Among these algorithms, XGBoost demonstrated the highest level of accuracy in lithofacies classification.

In another study by Kim [30], a pioneering approach was proposed for lithofacies classification in the challenging Austin Chalk and Eagle Ford formations, renowned for their suboptimal reservoir quality. The researchers introduced a CNN to tackle this classification task using conventional well logs, and remarkably, the CNN model outperformed the traditional ANN model. This research underscores the significance of harnessing cutting-edge methodologies like CNNs to significantly enhance the precision of lithofacies classification. An additional advantage of the CNN model is its reduced dependency on interpreted wireline logs, such as porosity, saturation, and brittleness, which mitigates the uncertainties accompanied by subjective interpretations due to manual intervention.

To tackle the complexities of lithofacies classification in a dynamic subsurface setting, a novel approach was introduced by Datta et al. [31]. This approach follows a multi-stage change detection process. It commences by detecting substantial variations in well logs, aligning these variations with lithofacies categories, optimizing the dataset by handling overrepresented classes, and finally applying the SVM for classification. Impressively, this method outperformed the traditional SVM algorithm.

#### IV. RESERVOIR CHARACTERIZATION

The procedure of objectively assigning reservoir attributes based on geological information and identifying uncertainties in geographical variability is referred to as reservoir characterisation [32].

Broadly, reservoir characterisation is performed during the exploration phase to assess the location and magnitude of possible oil reserves. Once it has been determined where and how many hydrocarbons are present in the reservoir, the oil field may be exploited to extract these reserves. Exploratory drilling often occurs in several separate wells during the first phase of this procedure. The objective of each well is to offer details on the features of the rock formation that surrounds the borehole and the types of hydrocarbon reserves that could be located there.

The objective of reservoir characterization is to obtain a deeper knowledge of reservoirs' physical and chemical features to make more knowledgeable choices about their development and exploitation, which affect the profitability of petroleum operations and their environmental impact. They help determine the best production methods to maximise output by indicating how reservoir fluid behaviour will change under various conditions. Reservoir characterisation aims to create a geological model that uses existing data to predict petrophysical properties across the oilfield [33]. Developing a precise image of a reservoir's characteristics may be challenging and time-consuming. Consequently,

there is a continual need to enhance automated reservoir characterization approaches.

#### V. PETROPHYSICAL PROPERTIES PREDICTION

It is essential to collect precise data on reservoir properties for reservoir characterization. The primary objective of reservoir characterization is to develop 3D representations of petrophysical characteristics. It comprises gathering data on petrophysical features, providing more insight into the fluid accumulation inside the rock formation. The most accurate method for estimating petrophysical properties is the laboratory-based method; however, it is expensive and time-consuming. Because of this limitation, there is only a limited number of samples accessible for certain wells, and these samples only cover a chosen number of depth intervals [34]. A significant number of samples are needed to accurately define a subsurface formation because of the complicated geological behaviours and spatial heterogeneity of reservoirs. Log-based approaches have been widely used to address this issue.

Actual samples of rocks were examined in a laboratory, and instrumental procedures that quantify physical qualities were used as data sources for petrophysical parameters [35]. These included core, seismic, and well logs. According to Xu et al. [36], petrophysical data can be regarded as big data as it meets the characteristic. Table 1 shows the petrophysical data. Machine learning has been widely used to predict petrophysical properties such as porosity, permeability, capillarity pressure, and water saturation. Machine learning eliminates the need for human processing and the geological complexities that traditional techniques must contend with, allowing for a significantly shorter processing time while maintaining exceptional quality and consistency in the output [37].

Significant subsurface parameters must be identified or evaluated; however, the most important factors are permeability and porosity. They are crucial indicators of the quality and financial feasibility of oil reservoirs. Porosity, a measurement of the proportion of open spaces or pores in a rock, is a crucial factor to consider when estimating the potential amount of hydrocarbons contained in a reservoir. The open spaces might serve as storage areas for hydrocarbons. Meanwhile, permeability is an important factor in characterizing how adjoined a rock's distinct open spaces are. Permeability measures the ability of hydrocarbons to flow up through the pores toward the surface where they may be taken out. It is impossible to obtain accurate solutions to many petroleum engineering issues without an accurate figure of permeability.

##### A. PERMEABILITY

A study by Huang et al. [38] investigated the application of an ANN to predict the permeability in an offshore gas field in eastern Canada. The authors proposed a back-propagation ANN using well log data from six wells. The proposed model surpassed conventional techniques such as multiple linear regression (MLR) and multiple nonlinear regression (MNLRL).

**TABLE 1. Common Petrophysical data and their attributes.**

Data	Source	Type	Depth-index	Dimension
Routine Core Analysis	Core	Numerical	Discrete	Low
X-Ray Diffraction (XRD) Mineralogy	Core	Numerical	Discrete	High
SEM	Core	Image	Discrete	High
Thin Sections	Core	Image	Discrete	High
Core Photo	Core	Image	Discrete	High
Capillary Pressure	Core	Array	Discrete	High
Electrical properties	Core	Numerical	Discrete	High
Relative Permeability	Core	Array	Discrete	High
Facies Description	Core	Text	Discrete	High
Mud logging	Mud Log	Numerical and Text	Continuous	Low
Conventional Logs	Log	Numerical	Continuous	Low
Dielectric Log	Log	Numerical	Continuous	Low
Nuclear Spectroscopy	Log	Array	Continuous	High
NMR Log	Log	Waveforms/maps	Continuous	High
Image Log	Log	Image	Continuous	High
Sonic Log	Log	Waveforms/maps	Continuous	High
Formation Testing (pressure build-up/drawdown)	Log	Array (pressure vs. time)	Discrete	High
Pressure Transient	Production	3-D	Discrete/Time-index	High
Seismic Attributes	Seismic	3-D	Continuous	High

Similarly, Helle et al. [39] supported this finding by predicting the porosity and permeability of the North Sea reservoir. The model also outperformed conventional methods. Likewise, Singh [40] employed ANN to estimate permeability from conventional well logs. The authors highlighted the technique's capacity to generate a constant good match between the projected and actual outputs. A study by Abdideh [41] predicted the permeability in an oilfield in Iran using a feed-forward back-propagation ANN technique. Utilizing well logs for prediction, the technique has advantages over MLR regarding prediction accuracy. The ANN model in Ben-Awuah and Padmanabhan [42] was developed to predict the permeability of a sandstone reservoir. However, only porosity was used as the model input. Using only three well logs features: mobility index, neutron porosity, and bulk density, Elkatatny et al. [43] constructed an empirical formula from an ANN to estimate the permeability in a heterogeneous carbonate reservoir. The suggested ANN model provides slightly lower accuracy than the Adaptive Neuro-Fuzzy Inference System (ANFIS) but is better than SVM, yet the model provides an empirical equation. However, Basbug and Karpyn [44] investigated the relationship between the permeability and porosity, specific surface area, and irreducible water saturation. The authors suggested using the ANN model to predict permeability. The proposed approach displayed acceptable levels of accuracy. A study by Irani and Nasimi [45] introduced evolving ANN to predict permeability. The model utilized

a Genetic Algorithm (GA) optimizer in ANN to search for the optimal parameters for the network. The authors noted that the proposed model provided a higher accuracy than the conventional ANN. After applying Principal Component Analysis (PCA) to extract relevant features from well logs, Bagheripour [46] constructed a CM consisting of MLP, Radial Basis Function (RBF), and Generalized Regression Neural Network (GRNN), utilizing GA to predict permeability. The proposed CM model produced better accuracy than the individual methods. In addition to GA, Matinkia et al. [47] examined Particle Swarm Optimization (PSO) and Social Skidriver (SSD) algorithm to predict permeability using MLP in the Fahlian Chahi Formation. The MLP-SSD hybrid provided the best accuracy after outlier removal and feature selection with Shapley Additive explanations (SHAP) were carried out. Also, Zhao et al. [48] utilized SHAP to visualize and explain their predictions using LR, SVM, BPNN, RF, KNN, GBDT and XGBoost algorithms. However, XGBoost provided the most accurate results in their predictions. Likewise, Liu and Liu [49] predicted the permeability in the Ordon Basin using a hybrid of PSO and XGBoost. The authors also utilized SHAP for feature selection and interpretation to make the model more explainable. The proposed model performed better than CNN, Long short-term memory (LSTM), and gated recurrent unit (GRU).

Although ANN has been shown to be effective for predicting permeability, they have the disadvantages of slow convergence and trapping at local minima. A study

by Tahmasebi and Hezarkhani [50] investigated a Modular Neural Network (MNN) to predict permeability. The MNN model comprises several interconnected neural networks that effectively decompose a large issue into smaller components. This enables quicker, simpler, and more accurate predictions. The suggested model outperformed the conventional neural network regarding prediction accuracy and performance. In a research conducted by Jamialahmadi and Javadpour [51], an RBF neural network was proposed to predict permeability from porosity. This model distinguishes itself from conventional neural networks because of its universal approximation and higher learning pace. Similarly, [52] proposed utilizing a GA as an optimizer inside an ANN to determine the best parameters to decrease time while achieving the greatest achievable performance. This technique was used to predict permeability separately in an Iranian reservoir based on geological zonation. However, Aifa et al. [53] investigated the efficiency of hybrid models for predicting the permeability and porosity using well logs. The authors suggested a neuro-fuzzy system that combines ANN and Fuzzy Logic (FL) to reap the advantages of both approaches while outperforming the methods individually. To overcome certain limitations of ANN, Saljooghi and Hezarkhani [54] introduced wavelet theory. The suggested technique utilizes various wavelets as activation functions to estimate permeability. The technique used well logs as input and showed superiority over conventional ANN. Meanwhile, Baziar and Tadayoni [55] compared the performance of the Co-Active Neuro-Fuzzy Inference System (CANFIS), MLP and SVM to estimate the permeability in a tight sandstone reservoir. CANFIS provided the best accuracy at the expense of slow computational speed. Using only porosity, specific surface area and irreducible water saturation, Kamali et al. [56] proposed using Group Method of Data Handling (GMDH) algorithm to predict permeability in carbonate reservoirs from Russia and Iran. The proposed algorithm was able to predict permeability accurately and outperform polynomial regression, Support Vector Regression (SVR) and Decision Tree (DT) when compared.

A study by Hamada and Elshafei [57] introduced Nuclear Magnetic Resonance (NMR) to complement conventional well logs to address the heterogeneity of gas sand reservoirs. NMR has been noted to offer lithology-independent quantitative porosity and a reliable estimate of the hydrocarbon potential. The authors applied forward-feed ANN to predict the porosity and permeability of a heterogeneous gas sand reservoir. According to the findings, predictions using NMR combined with conventional logs provide more accuracy than predictions using only conventional logs.

Some authors have conducted research using different machine learning techniques. A study conducted by El-Sebakhy et al. [58] applied a Functional Network (FN) technique to predict permeability in a carbonate reservoir. Using a polynomial basis, the FN model's predictive performance showed a better correlation than the ANN, ANFIS, and statistical regression, benefiting from the model's basic

architecture. Conversely, Olatunji et al. [59] explored extreme learning machines in predicting permeability in carbonate Middle Eastern reservoirs. The suggested method is superior to the ANN and SVM in performance, accuracy, and rapid learning speed. On the other hand, Gholami et al. [60] examined the Relevance Vector Regression (RVR) in the prediction of permeability in a carbonate reservoir using GA as an optimizer. When the accuracy of the proposed method was compared with that of SVM, it showed a modest advantage. In another study, Abdulraheem et al. [61] investigated the FL technique to predict the permeability in a Middle Eastern carbonate reservoir. The authors noticed the efficiency of subtractive clustering over the grid partitioning technique. The suggested technique showed excellent matching and proved effective for predicting the permeability. Furthermore, Wang et al. [62] optimized FL using Student-Newman-Keuls as a feature engineering technique. The proposed model outperformed the conventional technique without an optimizer.

In a study by Zhang et al. [63] compared the performance of MLP, SVR and MLR in the prediction of permeability in a heterogeneous tight gas sand reservoir. Porosity and well logs were used as inputs. MLP and SVR displayed high prediction accuracy, with SVR having a slightly higher correlation and MLP having a marginally lower error measure. In another study, Sheykhinasab et al. [64] proposed carbonate reservoir permeability prediction using the Least Square Support Vector Machine (LSSVM) and Multilayer Extreme Learning Machine (MELM) algorithms. The authors utilized the Cuckoo Optimization Algorithm (COA), PSO and GA to optimize the models. After the Tukey method was used for outlier removal, the hybrid of MELM and COA provided the most accurate results.

On the other hand, Anifowose et al. [65] utilized an ensemble machine learning paradigm to overcome a single hypothesis of conventional computational intelligence (CI) techniques and Hybrid Intelligent Systems (HIS) and the choice of CI/HIS model parameters. A study by Bhatt [66] attempted to predict porosity, permeability, fluid saturation and lithofacies in the Oseberg field using a bagging technique of committee machines. The author used wireline and measurement while drilling (MWD) logs for real-time prediction. The committee machines proved to exhibit superior performance over a single neural network.

However, Chen and Lin [67] used commonly employed empirical formulas in reservoir characterization to construct a novel ensemble model to calculate the permeability. The ensemble model used Wyllie and Rose [68], Coates and Dumanoir [69], and Schlumberger [70] empirical formulas to form a committee machine. The proposed method produced far more reliable predictions than individual methods and offered considerably greater generalization. However, [71] used empirical formulas and multiple regression in their committee machine. The ideal combination of weights was determined using GA. The authors predicted the permeability

of a carbonate reservoir in the Balal oil field using conventional well logs data. Similarly, the committee machine produced more accurate predictions than the individual methods. In Helmy's [72] ensemble model, it consists of SVM, ANN and ANFIS. Permeability was predicted in an oil field in the Middle East using well logs. This demonstrates that heterogeneous ensemble models may improve performance more than individual models, as seen in the accuracy and generalization. On the other hand, Anifowose et al. [73] used well logs from a Middle Eastern carbonate reservoir and employed three feature selection algorithms to make permeability predictions. The SVM and Type-2 Fuzzy Logic (T2FL) were trained using FN, DT, and Fuzzy Information Entropy (FIE) feature selection strategies. The FN-SVM hybrid approach performed very well compared to the other hybrid and standalone models. In contrast, an innovative approach put forth by Masroor et al. [74] introduces the Multiple-Input deep Residual Convolutional Neural Network (MIRes CNN) for predicting permeability in the Azadegan oil field, Iran. This unique technique simultaneously utilizes two distinct datasets: Numerical Well Logs (NWLs) and Graphical Feature Images (GFIs). The GFIs were generated by converting the 1D vector of NWLs to 2D matrices. While the NWL datasets are handled by a Single-Input deep Residual one Dimensional CNN (SIRes 1D-CNN), the GFIs are processed by a Single-Input deep Residual two Dimensional CNN (SIRes 2D-CNN). Comparative analysis demonstrated that this proposed approach outperformed SIRes 1D-CNN, SIRes 2D-CNN, GMDH, and RF methods.

Using forward feedback propagation neural network, Anifowose et al. [75], [76] formed an ensemble model to predict permeability and porosity. The cornerstone of a variety is neural networks with a varying optimum number of hidden neurons, with a randomized number of hidden neurons and with various learning algorithms. A study by Anifowose et al. [77] proposed an ensemble SVM model to predict porosity and permeability. The suggested model makes predictions based on various optimum regularization parameter values. A comparison of the model's performance against that of an SVM implemented using the bagging approach, a traditional SVM, and an ensemble of Decision Trees proved the superiority of the proposed model. A study by Anifowose et al. [78] suggested an ensemble of Extreme Learning Machines (ELM) to predict porosity and permeability. The proposed model utilizes an FN technique for advanced feature selection, which makes it a hybrid. The model performance surpassed the conventional ELM and Random forest. Otchere et al. [79] developed a hybrid model that utilized Random Forest and Lasso Regularisation feature selection technique combined with XGBoost to accurately predict water saturation and permeability. Based on the results, it was found that the suggested hybrid model outperformed both the conventional XGBoost model and the hybrid model that integrated PCA and XGBoost.

Even though newer well logging methods are more accurate than older ones, researchers have shown little interest in refining their algorithms. Although researchers have shown a limited interest in developing their algorithms, modern well logging methods have been demonstrated to offer greater accuracy than traditional ones. According to the literature review for the permeability prediction, Table 2 provides a thorough summary of the various machine learning approaches used, the input parameters included, and the reservoir location examined.

## B. POROSITY

Researchers have commonly used ANN to predict porosity in various formations [39], [57], [66], [80], [81], [82], [83], [84]. Using a back-propagation ANN, Helle et al. [39] predicted porosity and permeability in the North Sea. Using density, neutron porosity, sonic and gamma-ray, the authors could predict the porosity and permeability in Jurassic reservoirs with acceptable accuracy. A comparative study by Konate et al. [82] examined two ANN models to predict permeability in the Zhenjing oilfield. GRNN and feed-forward back propagation neural network (FFBP) were the models that were examined. The GRNN displayed superiority in prediction accuracy. Similarly, Zhang et al. [85] examined GRU neural network in prediction of porosity. The GRU provides a fast and demands less computational resources for the prediction. The proposed model included a Copula function as a correlation analysis (CA) for feature selection. Compared to standalone GRU, Recurrent neural network (RNN), and MLP models, the model's superiority has been shown. In another study, Hamada and Elshafei [57] developed a model that uses NMR logs to augment traditional well logs for gas sand reserves. The study found that predictions utilising NMR with traditional logs are more accurate than solely traditional logs.

Researchers have successfully hybridised ANN with other methodologies to circumvent the limitations inherent to ANN. In a comparative study, Zargari et al. [81] compared ANN and ANFIS to predict the porosity and permeability in an Iranian carbonate reservoir. The ANFIS provided better accuracy than the ANN model. The authors also acknowledged the potential of genetic algorithms for enhancing the prediction of ANN. Also, Elkatatny et al. [83] compare ANN, ANFIS and SVM. However, the authors noted that ANN provides better accuracy. Conversely, Nourani et al. [84] utilized Hand-held X-ray fluorescence (HH-XRF) as input for porosity prediction in a chalk reservoir. The authors relied on the speed and accuracy provided by the HH-XRF approach for geochemical characterization. The RF, ANN, GA-ANN, and GA-RF techniques were used to determine the most accurate prediction approach. However, the GA-RF offered the highest level of accuracy. However, Lim and Kim [80] utilized fuzzy logic for the input parameter selection between well logs before applying ANN for prediction. In another study, Ahmadi and Chen [86] applied

**TABLE 2. Summary of literature on the prediction of permeability using machine learning.**

References	Technique	Inputs Parameters	Field/Reservoir Source
[38]	ANN	SP, GR, RHOB, DT, NPHI, DRHO, latitude, longitude	Offshore gas field, Eastern Canada
[51]	RBF	core porosity	Asmari oil reservoir, south of Iran
[39]	ANN	RHOB, GR, NPHI and DT	North Sea case
[66]	CM	wireline and MWD logs	Oseberg field
[40]	ANN	GR, NPHI logs and RHOB	Uinta Basin, Southwest Utah field
[67]	CMEF	GR, LLD, LLS, MSFL, CAL, NPHI, RHOB and DT	Southern Taiwan
[61]	FL	CT, DRHO, DT, DR, MSFL, NPHI, PHIT, RHOB, RT, SWT	Middle Eastern carbonate reservoir
[44]	ANN	PHIT, SSA, and Swirr from core analysis	Carbonate reservoirs, USSR
[57]	ANN	GR, RHOB, NPHI, RT, NMR log data	Lower-Mesozoic reservoir
[52]	GA-ANN	SGR, RHOB, NP, PHIT, CT, SWT	Mansuri oil field, Iran
[71]	GA-Committee Machine	conventional well log data	Balal oil field, Persian Gulf
[45]	GA-ANN	GR, RHOB, NPHI, DT transit time, PHIT	Mansuri field
[50]	MNN, ANN	SGR, RT, SWT, PHIT and Secondary Porosity	Persian Gulf, Iranian offshore
[58]	FN, ANFIS, ANN, NR	CT, DRHO, DT, MSFL, NPHI, PHIT, RHOB, RT, SWT	Middle Eastern Carbonate reservoir
[41]	ANN, MLR	GR, DT, NPHI, RHOB, RT, SWT	Oil field, southwest of Iran
[59]	ANN, SVM and ELM	CT, DRHO, DT, MSFL, NPHI, PHIT, RHOB, RT and SWT	Middle Eastern reservoir
[65]	T2FL, SVM and ANN	NPHI, RHOB, CALI, GR, SWT, and seismic parameters	Middle Eastern carbonate reservoir
[76]	Ensemble ANN	GR, PHIT, RHOB, SWT, RT, MSFL, NPHI and CALI	Middle Eastern carbonate reservoir
[72]	Ensemble of SVM, ANN, ANFIS	MSFL, NPHI, PHIT, RHOB, SWT, CALI, CT, DRHO, GR, RT	Middle Eastern reservoir
[73]	SVM, FN, T2FLS, FN-T2FLS, FIE-SVM, FN-SVM, DT-SVM	GR, PHIT, RHOB, RT, MSFL, CALI, CT, NPHI, ILD, EL, Magnetic resonance	Middle Eastern carbonate reservoir
[53]	ANN, FL, NF	GR, RHOB, RT, NPHI, SWT	Hassi R'Mel field, Algeria
[60]	GA-RVR, GA-SVR	GR, DT, RHOB, NPHI, LLD, PEF and MSFL	Carbonate oil field, Kuzestan, Iranian
[46]	MLP, RBF, GRNN, CM-GA(MLP-RBF-GRNN)	DT, RHOB, NPHI, RHOB	-
[54]	ANN, WNN	RT, SWT, PHIT, SGR, Sand, Dol, shale and secondary porosity	Persian Gulf, Iranian offshore
[75], [77]	SVM, RF, SVM bagging, Stacked ensemble SVM	GR, PHIE, RHOB, SWT, RT, SFL, NPHI and CALI	Middle Eastern reservoir
[63]	SLR, MLR, MLP, and SVR	GR, AC, RT, RHOB, CNL, RT, RS and Coordinates(X, Y)	Chuanxi Depression, Sichuan Basin, China
[43]	ANN, SVM, and ANFIS	RHOB, NPHI, LLS, and LLD	Middle Eastern carbonate reservoir
[55]	CANFIS, MLP and SVM	GR, NPHI, RT, RHOB and DT	Measverde tight gas sandstones
[62]	FL	GR, SP, LLD, LLS, RHOB, AC and Elec	Mesozoic buried hill
[79]	RF-LR-XGBoost, PCA-XGBoost	CALI, DRHO, DT, GR, NPHI, PEF, RACEHM, RACELM, RD, RT, RHOB and ROP	Volve Oil Field, North Sea

**TABLE 2. (Continued.) Summary of literature on the prediction of permeability using machine learning.**

[48]	SHAP with LR, BPNN, SVM, RF, KNN, GBDT and XGBoost	GR, CAL, AC, DEN, CNC, RFOC, RILD, RILM, and SP	Wenchang A Sag
[49]	SHAP-PSO-XGBoost, CNN, LSTM, GRU	Por, CAL, GR, AC, RILD, SP, VSH, RFOC	Ordos Basin
[56]	GMDH, PR, SVR, DT	PHIT, SWT, SP	Carbonate Reservoir in Russia and Iran
[47]	MLP-SSD, MLP-PSO, MLP-GA	HCAL, CGR, PEF, NPHI, RHOB, RT, DTCO	Fahlian Chahi Formation
[64]	MELM-COA, MELM-PSO, MELM-GA, LSSVM-COA, LSSVM-PSO, LSSVM-GA, LSSVM, CNN	HCAL, SGR, PEF, NPHI, RHOB, RT, DTCO	Fahlian Formation, Iran
[74]	SIRes 1D-CNN, SIRes 2D-CNN, MIRes CNN, GMDH, RF	GFI, RHOB, NPHI, PEF, LLD, LLS, SGR, CGR, CAL, DCALI, and DRHO	Sarvak formation, Iran

**Abbreviations:** AC= Acoustic Log, ANN= Artificial Neural Network, BPNN= Backpropagation Neural Network, CAL= Caliper Log, CANFIS= Complex-Valued Adaptive Neuro-Fuzzy Inference System, CM= Committee Machine, CMEF= Committee Machine with Empirical Formulae, CNN= Convolutional Neural Network, CNC= Neutron Porosity Log, CNL= Compensated Neutron Log, COA= Cuckoo Optimization Algorithm, DCALI= Density Caliper Log, DEN= Density Log, DR= Deep Resistivity Log, DRHO= Bulk Density Log, DTCO= Compressional Sonic Log, DT= Sonic Log, ELM= Extreme Learning Machines, FL= Fuzzy Logic, GA= Genetic Algorithm, GBDT= Gradient Boosting Decision Trees, GMDH= Group Method of Data Handling, GR= Gamma Ray Log, GRNN= General Regression Neural Network, GRU= Gated Recurrent Unit, HCAL= High-Resolution Caliper Log, ILD= Deep Induction Log, KNN= k-Nearest Neighbors, LLD= Deep Laterolog Log, LLS= Shallow Laterolog Log, LSTM= Long Short-Term Memory, MWD= Measurement While Drilling, MSFL= Micro Spherically Focused Log, NPHI= Neutron Porosity Log, PEF= Photoelectric Effect Log, PCA= Principal Component Analysis, PHIE= Effective Porosity, PHIT= Porosity, PR= Polynomial Regression, RACEHM= Resistivity Log, RACELM= Resistivity Log, RBF= Radial Basis Function Network, RHOB= Bulk Density Log, RF= Random Forest, RILD= Deep Laterolog Resistivity Log, RILM= Shallow Laterolog Resistivity Log, RT= Resistivity Log, RVR= Relevance Vector Regression, RT= Resistivity Log, SFL= Shallow Micro Spherically Focused Log, SGR= Shale-Gas Ratio Log, SHAP= Shapley Additive Explanations, SIRes 1D-CNN= Single-Input deep Residual 1D-CNN, SIRes 2D-CNN= Single-Input deep Residual 2D-CNN, SLR= Simple Linear Regression, SSA= Specific Surface Area, SVM= Support Vector Machine, SVM= Support Vector Regression, Swirr= Irreducible Water Saturation, SWT= Water Saturation, T2FL= Two-Step Fuzzy Logic, T2FLS= Type-2 Fuzzy Logic System, VSH= Volume of Shale Log, WNN= Wavelet Neural Network, XGBoost= Extreme Gradient Boosting

an Imperialist Competitive Algorithm (ICA) and a hybrid GA and PSO (HGAPSO) to predict porosity using an ANN. The author also applied HGAPSO optimization to the LSSVM for porosity prediction. The models were compared with standalone ANN and fuzzy decision trees (FDT). However, the HGAPSO-LSSVM model provided the high accuracy. Furthermore, Sun et al. [87] suggested optimizing the Elman neural network with a Whale Optimization Algorithm (WOA) to predict porosity in oil wells in Western China. Compared to the standalone Elman and BP algorithms, the WOA-Elman algorithm provided better accuracy. A study by Wang and Cao [88] proposed a prediction of porosity using a deep learning method called an integrated neural network. The suggested approach, combining a 1-dimensional CNN with bidirectional GRU, demonstrated higher accuracy than the biGRU, GRU, LSTM, RNN and MLR.

Other machine learning techniques have also been used to predict porosity. A study by Al-Anazi and Gates [89] investigated the SVR technique to estimate the porosity. The proposed model proved superior to the MLP, GRNN and Radial Basis Function Neural Network (RBFNN) in terms of accuracy and robustness. However, the SVR robustness is subject to kernel function selection. The advantage comes with the burden of using far more computational resources than various alternative approaches. However, Anifowose et al. [73] applied three feature selection techniques for porosity prediction using laboratory measurements from the Northern Marion Oilfield. FN, DT, and Fuzzy Information Entropy(FIE) feature selection techniques were applied to the SVM and T2FL. The FN-SVM hybrid technique proved outstanding among the alternative hybrid and standalone models. Furthermore, Ahmadi et al. [90] employed GA's optimization ability to perform predictions using FL and

LSSVM. The suggested models predicted the porosity and permeability of wells from Northern Persian Gulf oilfields. GA-LSSVM provided slightly better accuracy than the alternative method. Also, Zhong and Carr [91] investigated a hybrid SVM model with a mixed kernel function (MKF). The model was optimized using particle swarm optimization (PSO) to improve its predictive capabilities. Regarding accuracy, the proposed method outperformed the conventional SVM, LSSVM, ANN, and RBF. In a separate study, Andersen et al. [92] undertook an optimization of the LSSVM to predict porosity and water saturation in the Varg field located in Norway. The authors explored predictive models using various combinations of well logs. Interestingly, their findings highlighted that the most accurate predictions were achieved when focusing on porosity and utilizing only three specific well logs: density, deep resistivity and gamma-ray logs. Moreover, their research indicated that incorporating additional well logs yielded no noticeable enhancements in the model's predictive performance. In addition, Anifowose et al. [77] presented an ensemble model using SVM. The proposed model offers predictions based on several optimal regularisation parameter values. A comparison of the performance of the proposed model with that of an SVM implemented using the bagging technique, a standard SVM, and an ensemble of DT demonstrated its superiority. In another study, Tariq et al. [93] compared deep neural network (DNN), DT, RF KNN, XGBoost, and AdaBoost for predicting NMR porosity using conventional well logs. Based on the outcome, it was found that DNN, RF and XGBoost demonstrated superior levels of accuracy. The experimental results strongly indicate that employing DNN, RF, or XGBoost can significantly enhance the accuracy of predictions.

On the other hand, Haqqi et al. [94] suggested predicting porosity in the Damar field, Indonesia, using the XGBoost algorithm optimized with the GridSearchCV(GS) technique. However, Pan et al. [95] proposed predicting the porosity using an optimized XGBoost model with GS and GA. The presence of two different optimization strategies benefits the model, as shown by its accuracy. The suggested model outperforms the alternatives when examined with a GS optimisation model alone, followed by LR, SVR, RF, and XGBoost. According to the literature review for the porosity prediction, Table 3 provides a thorough summary of the various machine learning approaches used, the input parameters included, and the reservoir location examined.

### C. WATER SATURATION PREDICTION

Water saturation is another vital reservoir property indicating the water portion present in certain pore spaces. It aids in calculations of perforation depth for offshore and onshore hydrocarbon-producing sites [96]. It is necessary for the appropriate computation of hydrocarbon volume. Over the last few decades, various empirical methods for predicting

water saturation have been introduced using petrophysical data from logs, including resistivity, sonic, density, and neutron porosity. The pioneering empirical model for predicting saturation was the Archie [97] model for clean sandstone reservoirs. Several researchers have attempted to derive the relationship between water saturation and well log data to predict water saturation in different formations [98], [99], [100], [101], [102]. However, these approaches are constrained by their formation and are only applicable in restricted lithologies. These models lack generalization and cannot be applied universally. Furthermore, the parameters associated with each model have their underlying uncertainties, which may lead to misinterpreted outcomes. Therefore, machine learning techniques have been widely used to predict water saturation.

ANN and FL are examples of popular Artificial Intelligence (AI) techniques used to predict water saturation. Among the many different machine learning approaches, ANN has the widest range of potential applications and has been shown to be successful in various contexts. Several ANN models have been successfully applied to core data and well logs. The earliest was Helle and Bhatt [103], which proposed a committee neural network that utilized sonic, density, neutron porosity and resistivity logs as inputs. Subsequently, Shokir [104] implemented an ANN model that included the self-potential log (SP log) to the input features. The model's superiority was proved by comparing the water saturation predictions generated by ANN with those generated by conventional petrophysical analysis. On the other hand, Kamalyar [105]'s model solely considered the porosity and permeability from the core as well as the height above the free water level. Similarly, Al-Bulushi et al. [106] proposed an ANN trained using a resilient back-propagation learning algorithm. The authors also investigated the effect of several different well log parameters that were the model's inputs using a feature ranking approach carried out on the well logs. The proposed model was used to predict water saturation, providing better accuracy than the statistical approach. In addition to this, Mardi et al. [107] also out the idea of using an ANN model to predict not only water saturation but also cementation and the saturation exponent in two carbonate reserves located in Iran. The model used by the authors included not just well log measurement but also core porosity. Water saturation, porosity and permeability in the Niger Delta region were predicted by Okon et al. [108] using a feed-forward back-propagation ANN. The proposed model included feature ranking and achieved high accuracy. In Al-Bulushi et al. [109]'s study, density, neutron, resistivity, and photo-electric wireline logs were selected as input features to construct a model using an ANN technique to predict water saturation. According to Nyein et al. [110], the superiority of the ANN model over the conventional models in predicting the water saturation and porosity in a shaly sandstone reservoir was reported. The core data from the two wells exhibited an excellent fit to the suggested model, which demonstrated good matching.

**TABLE 3.** Summary of literature on the prediction of porosity using machine learning.

References	Technique	Inputs parameter	Field/reservoir source
[39]	ANN	DEN, DT, RT	-
[80]	FL-ANN	NPHI, CAL, LLD, LLS, RHOB, and SP	-
[89]	SVM	GR, CNL, DT, DEN	-
[39]	ANN	NMR logs and GR, RHOB, CNL, Deep and Shallow RL	Gas sand reservoir
[57]	ANN	GR, RHOB, CNL, Deep and Shallow RL	Gas sand reservoir
[81]	ANN	DT, NPHI, RHOB, CT, and SGR	Carbonate reservoir, Southern Iran
[73]	SVM, FN, T2FLS, FN-T2FLS, FIE-SVM, FN-SVM, DT-SVM	petrographic measurements - top interval grain density, grain volume, length and diameter	Northern Marion Platform of North America
[90]	GA-FL, GA-LSSVM	DT, DEN, CNL, PHIT	Northern Persian Gulf oil-fields
[78]	SVM, RF, SVM bagging, Stacked Ensembles	petrographic measurements - top interval, grain density, grain volume, length and diameter	Northern Marion Platform of North America
[82]	ANN	RHOB, CNL, AC, and NPHI	-
[43]	ANN, ANFIS, SVM	DT, RHOB, and NPHI	Carbonate reservoir
[86]	FDT, ANN, ICA-ANN, HGAPSO-LSSVM	DT, RHOB, PHIT, NPHI	Iranian oil fields
[91]	PSO-MKF-SVM	GR, DEN, Shale content, Slope of GR, Slope of density	Jacksonburg-Stringtown oil field
[85]	GRU NN, CA-GGRU, MLP, RNN		Ordos basin
[84]	RF, ANN, GA-RF, GA-ANN	Hand-held X-ray fluorescence (HH-XRF)	-
[88]	INN, BiGRU, GRU, LSTM, RNN, MLR	GR, DEN, CNL, DT, NPHI,	Western Sichuan Basin
[87]	Elman, WOA-Elman, BP	PE, DEN, M2R1, AC, GR, R25, R4 and CNL	Oilfield in Western China
[95]	LR, SVR, RF, XGBoost, GS-XGBoost, GS-GA-XGBoost	AC, CAL, CNL, DEN, GR and array induction resistivity(AT90)	Oilfield in Northern Shaanxi, China
[92]	LSSVM-PSO	RHO, RD, GR	Varg field, Norway
[94]	GS-XGBoost	ILD, NPHI, RHOB, GR, rock formation	Damar field, Indonesia
[93]	DNN, DT, RF, KNN, Ad-aBoost, XGBoost	GR, CALI, NPHI, PE, RHOB	-

**Abbreviations:** AC= Caliper Log, **AdaBoost**= Adaptive Boosting, **ANN**= Artificial Neural Network, **ANFIS**= Adaptive Neuro-Fuzzy Inference System, **BP**= Backpropagation, **CA**= Correlation Analysis, **CAL**= Caliper Log, **CNL**= Neutron Density Log, **DNN**= Deep Neural Network, **DT**= Sonic Log, **DEN**= Density Log, **FL**= Fuzzy Logic, **FN**= Functional Network, **GA**= Genetic Algorithm, **GB**= Gradient Boosting, **GR**= Gamma Ray Log, **GRU**= Gated Recurrent Unit, **GS**= Genetic Search, **HGAPSO**= Hybrid Particle Swarm Optimization and Genetic Algorithm, **ICA**= Imperialist Competitive Algorithm, **ILD**= Deep Induction Log, **LLD**= Deep Laterolog Log, **LLS**= Shallow Laterolog Log, **LR**= Lasso Regularization, **LSSVM**= Least Square Support Vector Machine, **MKF**= Mixed Kernels Function, **MLP**= Multilayer Perceptrons, **NPHI**= Neutron Porosity Log, **PCA**= Principal Component Analysis, **PE**= Photoelectric Effect, **PSO**= Particle Swarm Optimization, **RHO**= Resistivity Log, **RD**= Deep Resistivity Log, **RF**= Random Forest, **RNN**= Recurrent Neural Network, **SVM**= Support Vector Machine, **SVR**= Support Vector Regression, **T2FLS**= Type-2 Fuzzy Logic System, **WOA**= Whale Optimization Algorithm, **XGBoost**= eXtreme Gradient Boosting.

Another study by Kenari and Mashohor [111] formulated an ANFIS by combining ANN and fuzzy logic to estimate the water saturation in a carbonate Iranian field. The model is superior to the conventional ANN as it can deliver more

accuracy, robustness, and generalisation than each of the separate components. In addition, Ibrahim et al. [112] compared empirical equations with ANN and ANFIS to predict water saturation. The ANFIS slightly outperformed the ANN in

the prediction outcome but was significantly better than the empirical formulae. Additionally, Khan et al. [113] compared ANN and ANFIS in a carbonate reservoir in the Middle East. The results showed that ANFIS provided slightly better output accuracy than ANN. Meanwhile, ANN and FL were compared by Bageri et al. [114] in a carbonate reservoir in the Middle East. The output suggests that the FL model offers higher accuracy than the ANN model.

Conversely, Amiri et al. [115] optimized their ANN model using an Imperialist Competitive Algorithm (ICA) in an unconventional reservoir. Furthermore, the authors noticed the impact of outliers, which significantly improved the prediction outcome when detected and deleted when appropriate. In another study, Gholanlo et al. [116] proposed the concept of using a radial basis neural network to predict water saturation in the carbonate Sarvak Formation in Iran. Compared to other neural network models, the advantages of the RBF model include its straightforward structure and ability to acquire knowledge quickly.

However, Adeniran et al. [34] reported the efficiency of FN in predicting water saturation and reservoir porosity using well logs. This model has been noted to produce a speedy and unique solution that surpasses neural networks. Also, Tariq et al. [117] suggested an FN model to predict the water saturation. The model's accuracy was improved by trying many optimization algorithms, such as Differential Evolution, PSO, and Covariance Matrix Adaptation Evolution Strategy (CMAES) to develop the most accurate version. The PSO proved to be the best choice among them. Additionally, Andersen et al. [92] conducted an optimization of the Least Squares Support Vector Machine (LSSVM) for predicting porosity and water saturation in the Varg field, Norway. Their investigation involved predicting using different sets of well logs. Surprisingly, the best results were achieved when predicting water saturation using only four logs: medium resistivity, gamma ray, adjusted caliper, and self-potential logs. Interestingly, their study revealed that the inclusion of additional logs did not lead to any improvement in predictive performance.

SVM is yet another alternative technique to machine learning that has been presented for predicting water saturation by [96], [118], and [119]. According to Mollajan et al. [118], the model outperformed the ANN in terms of the accuracy of its predictions. Furthermore, Miah et al. [96] examined another version of SVM, which used least squares as its kernel function called least-squares support vector machine (LS-SVM). The authors also considered the significance of feature ranking because it reduces the model's time and complexity by considering only the most important input characteristics. Their proposed LS-SVM surpassed the ANN in terms of predictive accuracy.

Baziar et al. [120] compared the performance of an SVM, ANN, Random forest and gradient boosting to predict water saturation using a small data set in a sandstone reservoir. Although the authors reported the reliability of all the various techniques used, SVM was

noted to provide the best performance. In another study, Hadavimoghaddam et al. [121] compared the accuracy of various boosting algorithms, namely XGBoost, LightGBM, AdaBoost, CatBoost and Super Learner, to predict water saturation in a sandstone reservoir in the Russian Federation. Of all the options, XGBoost proved to be the most precise. Nevertheless, the accuracy was only slightly better than that of the Super Learner.

Otchere et al. [79] constructed a hybrid model consisting of an ensemble model of Random Forest and Lasso Regularisation as the feature selection technique and XGBoost as a predictor to predict water saturation and permeability. The suggested hybrid model was better than the traditional XGBoost model and a hybrid model that included PCA and XGBoost. According to the literature review for water saturation prediction, Table 4 thoroughly summarises the various machine learning approaches used, the input parameters included, and the reservoir location examined.

## VI. DISCUSSION

Machine learning has seen remarkable growth in oil and gas exploration. This can be attributed to its ability to address various challenges in the industry, such as seismic data processing, lithofacies classification and reservoir characterization.

Machine learning models have several benefits over traditional oil and gas exploration approaches, derived from empirical and semi-empirical models for estimating reservoir parameters. Machine learning models can discover insights from the well logs that traditional models have overlooked by capturing the high-dimensional complicated interactions and nonlinear behaviours among the well log parameters. Furthermore, they yield remarkably accurate results using significantly less time and resources than traditional methods [112]. The benefits of machine learning cannot be overstated because it is evident that they may dramatically decrease the time required for seismic data processing, lithofacies classification and reservoir characterization. Similarly, as a result, the amount of labour and resources needed to address problems in the industry is decreased [13], [25], [63].

However, there are limits to what can be accomplished with every method, and machine learning is no exception. Despite significant progress in tackling linear, nonlinear, and complicated problems, including classification, regression, and prediction, several downsides exist. The commonly used MLP is sluggish to train, prone to becoming trapped in local minima and requires a lot of trial and error to determine the ideal topology. Additionally, it demands a greater quantity of data than its counterpart models.

Furthermore, there is a direct correlation between the quality of the data used to train a machine learning algorithm and the performance of the algorithm itself. The data quality used to train the model directly affects how accurate it is [122]. This is frequently referred to as the GIGO principle (garbage in, garbage out). This indicates that a poor

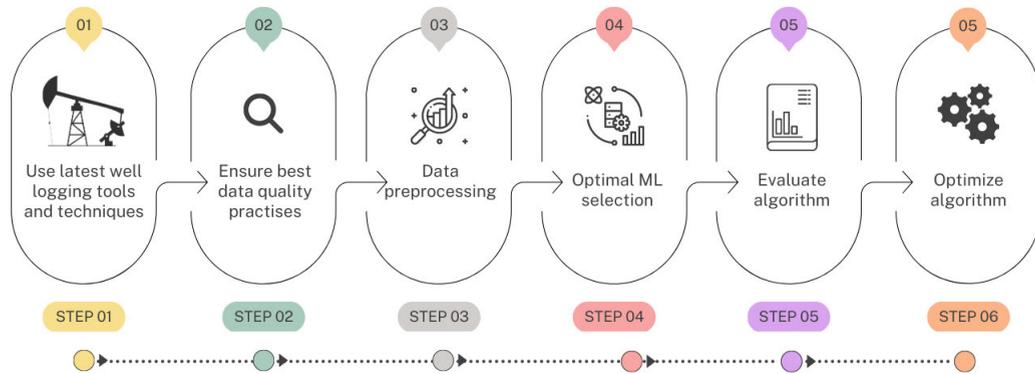
TABLE 4. Summary of literature on the prediction of water saturation using machine learning.

References	Technique	Inputs parameter	Field/reservoir source
[104]	ANN	SP, DT, RHOB, NPHI, resistivity log	Egyptian shale field
[119]	SVM	RHOB and resistivity logs	Gulf of Mexico
[106]	ANN	GR, NPHI, RHOB, resistivity wireline data	Sandstone, Haradh formation, Oman
[34]	FN	NPHI, RHOB, GR, Rt, and PEF	Middle Eastern
[105]	ANN	Porosity and K from core, and h*	Iran Carbonate
[107]	ANN	RT, NPHI, RHOB, DT, and core porosity	Carbonate, Azadegan oilfield, Sarvak formation, Iran
[109]	ANN	RHOB, NPHI, resistivity, and PE	Sandstone, Gharif and Hardh formations of Oman
[111]	ANN, FL, ANFIS	SP, RT, NPHI, RHOB, PE, and effective PHI	Carbonate, Khark oilfields, Iran
[118]	ANN, SVR	RT, NPHI, RHOB, and DT	Carbonate, Sarvak formation, Iran
[115]	ICA-ANN	GR, RT, NPHI, RHOB, and effective NPHI DP	-
[114]	ANN, FL	Resistivity logs and core data	Middle Eastern carbonate reservoir
[116]	RBF ANN	NPHI, DT, RHOB, and core data	Carbonate, Sarvak formation, Iran
[110]	ANN	GR, LLD, RHOB, NPHI, PEF	Upper Cretaceous shaly sand formation in Western Desert, Egypt
[113]	ANN, ANFIS	GR, RT, R <sub>xo</sub> , NPHI, RHOB, and CALI log	Middle Eastern carbonate reservoir
[120]	SVM, ANN, RF, GB	GR, NPHI, ILD, RHOB, DT	Mesaverde tight gas sandstones, Uinta Basin
[96]	LSSVM, ANN	DT, GR, RT, RB, PE, NP	-
[117]	PSO-FN, DE-FN, CMAES-FN	GR, RHOB, NPHI, 15FR, LLS, and LLD	Middle eastern carbonate reservoir
[108]	ANN	GR, DL, RL	Niger Delta region
[121]	XGBoost, LightGBM, AdaBoost, CatBoost and Super Learner	GR, NPHI, RHOB, DT, PEF	Russian Federation Sandstone reservoir
[79]	RF-LR-XGBoost, PCA-XGBoost	CALI, DRHO, DT, GR, NPHI, PEF, RACEHM, RACELM, RD, RT, RHOB and ROP	Volve Oil Field, North Sea
[112]	ANN, ANFIS	GR, NPHI, Rt	Tight Gas Sandstone Formation
[92]	LSSVM-PSO	Adjusted CAL, RM, SP, GR	Varg field, Norway

**Abbreviations:** **AdaBoost**= Adaptive Boosting, **ANN**= Artificial Neural Network, **ANFIS**= Adaptive Neuro-Fuzzy Inference System, **CALI**= Caliper Log, **CatBoost**= Categorical Boosting, **CMAES**= Covariance Matrix Adaptation Evolution Strategy, **DE**= Differential Evolution, **DT**= Sonic Log, **FL**= Fuzzy Logic, **FN**= Functional Network, **GB**= Gradient Boosting, **GR**= Gamma Ray Log, **ICA**= Imperialist Competitive Algorithm, **ILD**= Deep Induction Log, **LightGBM**= Light Gradient-boosting Machine, **LR**= Lasso Regularisation, **LLD**= Deep Laterolog Log, **LLS**= Shallow Laterolog Log, **NPHI**= Neutron Porosity Log, **PCA**= Principal Component Analysis, **PEF**= Photoelectric Effect Log, **PSO**= Particle Swarm Optimization, **RACEHM**= Resistivity Log, **RACELM**= Resistivity Log, **RBF**= Radial Basis Function, **RF**= Random Forest, **RT**= Resistivity Log, **SVM**= Support Vector Machine, **SVR**= Support Vector Regression, **XGBoost**= Extreme Gradient Boosting.

representation of the challenge inadequately represents the situation’s dynamics, which is necessary to discover how to translate instances of inputs into outcomes. The original data may have been compressed into nonlinear correlations

revealed only after extensive data preprocessing. The data may also be flawed for various reasons, such as values that are out of range, contradictory information or minor random changes in observations. As a result, substantial



**FIGURE 3.** Road-map for optimal accuracy of machine learning techniques in oil and gas exploration.

data preparation must be conducted to capture the intricate interaction of variables that might be discovered across data sources in the upstream oil and gas industry.

Despite their effectiveness, individual machine learning models are not sufficiently resilient to address complicated issues and deal with uncertainties in the oil and gas sector. Researchers have recently focused on using ensembles and hybrid machine learning approaches to overcome this issue. This is shown by the growing number of recent articles on using ensembles and hybrid machine learning techniques for seismic data processing, facies and lithofacies classification and reservoir characterization. Despite this, a considerable amount of work still has to be carried out to standardise the techniques for ensemble integration.

Hybrid machine learning was used to supplement the individual models with the strengths of others. Hybrid machine learning combines diverse computations or processes from multiple models, all intended to improve one another. Various basic models collaborate to complete and strengthen one another to produce improved outcomes compared to their single model equivalents. Optimization algorithms such as GA may enhance models by selecting the best hyperparameters. Dimensional reduction algorithms, such as PCA, may decrease model complexity while simultaneously removing noise from data. There are various hybridization options to explore for improving single machine learning models capable of addressing the complex challenges of the oil and gas sector.

A committee machine was used to develop the neural network further. A more accurate, robust, and better capacity to make generalisations is achieved by combining the expertise of several experts rather than focusing solely on the superior expert. This is due to the fact that the generalization of individual members is not unique. Expert pruning may circumvent the extra resource restrictions imposed by the committee machine.

In addition, ensemble learning has been researched further to enhance the performance of individual machine learning models to address complicated problems in oil and gas exploration. Ensembles can integrate various outcomes, including multiple learning techniques, conflicting interpretations of

data, randomly sampled data considerations, multiple model structures, and other well-defined properties of interest. Because ensemble models can manage numerous hypotheses simultaneously, they can assist in overcoming the high degree of uncertainty present in reservoir attributes and model-tuning variables. This enables more reliable and accurate outcomes and provides overall conclusions with the least chance of error and ambiguity. Ensemble learning can manage the synthesis of highly dimensional and multi-modal data, such as those in the oil and gas sector. Ensemble learning has endless opportunities to be examined and analyzed to achieve potentiality and enhanced performance.

Machine learning can significantly change the decisions made by oil and gas industry specialists. Machine learning is expected to become increasingly important in the oil and gas sectors in the future years [20]. Nevertheless, researchers still face difficulties obtaining data from laboratories and fields, which is an obstacle to improving the literature.

As oil and gas exploration continues to generate massive amounts of data, it is becoming more important to create, improve, and incorporate big data management methods in the field of AI. Utilizing the available data to its fullest potential is a current focus, and it will likely remain in the future. To achieve optimization, one must make use of AI's formidable resources.

The road map shown in Fig. 3 are processes critical for achieving optimal accuracy in applying machine learning in oil and gas exploration. The first stage involves collecting high-quality data about the reservoir using the most recent well logging instruments and methodologies. These technologies may offer a wide variety of measurements that can be used to determine the geological, geophysical and characterization aspects of the reservoir. The second stage is to verify that the highest data quality methods are followed. Comprehensive data validation, cleaning, and normalisation are required to ensure the data is correct and dependable. The accuracy and efficacy of the machine learning algorithm are affected by the data quality employed for modelling. Data preparation is the third phase. This process includes selecting important data characteristics, scaling, and translating the data into a format suitable for

machine learning algorithms. It is critical to identify and eliminate features that are irrelevant to the problem at hand. The fourth step is choosing the appropriate machine learning algorithm for the data and task. Machine learning methods such as classification, regression, hybrid, and ensembles may be employed. The specific problem and dataset will determine the algorithm used. The fifth stage assesses the machine learning algorithm using several metrics such as accuracy, precision, mean absolute error, mean squared error and R-squared. This stage aids in determining the correctness of the model and identifying any improvement areas. The last phase is to increase the accuracy and performance of the machine learning algorithm. This may be accomplished by altering the hyperparameters or using a new method entirely. The objective was to obtain the highest possible prediction accuracy using the supplied data. Overall, the procedures shown in the figure are a good starting point for academics and practitioners interested in applying machine learning to predict reservoir properties in the oil and gas exploration stage.

## VII. CHALLENGES

The oil and petrol sector generates massive quantities of data through exploration, drilling, production, and refining operations, making it one of the most data-intensive industries in the world. The advantages of machine learning in the industry cannot be understood, as it can boost efficiency, lower costs, and increase safety. Nonetheless, some significant technological problems must be addressed to leverage the potential of machine learning in the exploration stage of the industry. These are described below.

### A. DATA ISSUE

#### 1) DATA AVAILABILITY

The lack of readily available high-quality data is a significant barrier to the widespread use of machine learning in the oil and gas industry. The oil and gas sector produces huge volumes of data, yet a lot of it is unstructured, dispersed, and difficult to access [123], [124]. This is a serious concern for machine learning algorithms because they function best when fed with massive volumes of high-quality data [125].

The exploratory stage for oil and gas contributes to the lack of data in the industry. During early exploration, limited data is common due to the difficulty of drilling wells in extreme conditions such as the deep sea or Arctic. This makes data collection and transmission from such areas laborious and expensive. Utilizing the obtained data in machine learning applications might be difficult if they are limited, incomplete, inconsistent, or of poor quality [1].

#### 2) DATA PREPROCESSING

Poor quality gives rise to a data preprocessing challenge due to the complexity of the data required for machine learning model training [126]. These data may include seismic surveys, drilling data, well logs, production data,

and other geophysical data of varying quality and format geophysical data. There might be a substantial number of redundancies, inconsistencies and missing values in the data, requiring extensive cleaning and standardization before the data can be useful.

Moreover, combining these diverse data sources results in a substantial volume of data, presenting challenges associated with high dimensionality due to the multitude of attributes measured at different depths and locations. Additionally, the inherently uncertain geological conditions contribute to further subsurface data uncertainties arising from measurement and calibration error, processing, interpolation, and extrapolation.

Reservoirs exhibit geological features across different scales, from microscopic pore-scale structures to macroscopic field-scale structures. Integrating data collected at various scales is crucial for developing accurate and comprehensive reservoir models. Exploration activities often involve spatial data, such as geological maps, seismic surveys, and satellite imagery, introducing unique challenges in data integration, feature engineering, and computational demands. Temporal information present in some exploration datasets, documenting historical changes in geology or environmental factors, adds another layer of complexity and uncertainties, requiring specialized techniques like time series analysis and data fusion for meaningful insights.

Preprocessing is significantly more challenging in carbonate reservoirs owing to their severe heterogeneity and complicated pore structure composed of matrix porosity, vugs, fractures, and other geological features [43]. This leads to a weak porosity-permeability relationship. Because of this, the permeability prediction using the NMR log becomes more difficult as it relies heavily on the correlation between porosity and permeability. Furthermore, outliers and anomalies could be present in the data, reducing the accuracy of the machine learning models.

#### 3) DATA FRAGMENTATION AND ACCESS RESTRICTIONS

The fragmented structure of the sector is another source of data scarcity [124]. Many firms, contractors, and service providers are engaged in exploration and production operations in the oil and gas sector, making it a highly decentralized industry. Data silos and restricted access result from this fragmentation, which makes it difficult to transfer data across various entities.

Lastly, legal and privacy concerns restrict data access in the oil and petrol industry. Because of the potentially sensitive nature of the data gathered during exploration and production, stringent rules limit its collection, use, and dissemination.

#### 4) ADDRESSING CHALLENGES

Potential solutions can be applied to overcome these challenges. One answer is that stakeholders in the sector should work together and share information. Developing common data standards, publishing data in public repositories, and

teaming up with academic institutions to build data-sharing infrastructure are viable options.

Data augmentation is an alternative approach when new information is added to preexisting data. This may entail generating synthetic data using simulation tools, augmenting seismic imagery with computer vision methods, or reusing data from other sources via transfer learning. Numerous methods, such as flipping, cropping, rotating, and adding noise to the original data, are used to create additional training data from preexisting data sets. This can be applied to image data types such as SEM, thin sections, cores and seismic images to increase the quality and quantity. Methods such as downsampling, upsampling interpolation, extrapolation, and smoothing can be implemented on the well logs.

Efforts should be undertaken to obtain additional data using novel techniques to increase data accessibility. Data from inaccessible areas can be gathered using remote sensing technology such as drones and satellite photos. Drilling and production data can be collected in real-time using modern sensor technology.

Enhancing the quality of current data is another way to address the issue of data scarcity. Investing in data preprocessing methods can help ensure sufficient data quality. This may involve data cleaning, normalization and transformation. A quality control approach can also be utilized to ensure adequate data quality for machine learning applications. This might include establishing uniform guidelines for data collection and conducting various data checks for consistency and validation. Dimensionality reduction techniques can be used to retain essential information while reducing the number of features. Feature selection methods can also be applied to identify and keep the most relevant attributes.

Furthermore, multiscale modelling techniques that consider both microscopic and macroscopic features can be utilized. This involves adapting algorithms to handle data at different scales and integrating diverse datasets for a comprehensive reservoir model. Also, uncertainty quantification techniques can be integrated into data preprocessing, which can help model and manage uncertainties, providing a more robust representation of geological conditions. Specialized techniques like spatial data integration, feature engineering, and computational methods to handle the unique challenges of spatial data can be implemented. For temporal data, employ time series analysis and data fusion techniques to extract meaningful insights from historical changes.

Models specifically designed for carbonate reservoirs should take into account the distinct characteristics of matrix porosity, vugs, fractures, and other features. Exploring advanced machine learning techniques that can handle weak porosity-permeability relationships would be beneficial. Additionally, it is important to carefully examine the implementation of outlier detection methods, as some unique subsurface structural features may be viewed as outliers, which can improve the model's accuracy.

Comprehensive methodologies capable of handling uncertainties, heterogeneity and complex structures must be

developed. The future requires robust, adaptable machine learning frameworks to handle data quality, uncertainties and limitation challenges. This will allow the establishment of a robust input-output relationship. Advanced machine learning approaches, such as feature selection, dimensionality reduction, and appropriate regularisation, may be used to capture complicated data correlations and improve forecast accuracy.

Research can focus on improving interpolation techniques to make predictions more accurate and robust, especially in areas with sparse data. Advanced spatial statistics and machine learning methods like Gaussian processes can be explored. Developing standardized data formats, ontologies, and metadata standards for geospatial data can aid in data integration. Automated tools for data harmonization can be created. Techniques for effective data fusion of temporal and spatial data can be developed. This can involve research in spatiotemporal databases and GIS (Geographic Information System) applications.

Furthermore, engaging with regulatory organizations to set data exchange and utilization rules greatly increases regulatory compliance. Methods for doing so include data-sharing agreements and data anonymization.

Future research can focus on novel approaches to effectively address the challenge of data scarcity. Improved information extraction may be possible by developing innovative data mining algorithms to handle vast and complicated data sets. Similarly, Predictive abilities can be improved by developing new machine learning algorithms optimised for learning from small and noisy data sets. In addition, developments in data fusion techniques have enabled data to be integrated from various sources more efficiently. Lastly, improved machine learning techniques for data discovery in limited data can be explored to identify new patterns and insights in limited data.

When addressing the problem of data issues in the oil and gas industry for machine learning purposes, a hybrid approach is most likely to provide the best results. A holistic approach involving data cleaning, dimensionality reduction, uncertainty management, multiscale modelling, and specialized techniques for spatial and temporal data is essential to address the data preprocessing challenges in the oil and gas exploration sector. Tailoring solutions to specific geological conditions, such as those in carbonate reservoirs, further enhances the effectiveness of machine learning models. Furthermore, the oil and gas sector can realize the full benefits of machine learning in exploration and production if its members work together to enhance data sharing, collection, and quality. Future research directions should also be embraced in the development of innovative approaches.

## **B. TRANSPARENCY AND INTERPRETABILITY OF MODELS**

### **1) TRANSPARENCY OF MODELS**

Identifying how machine learning models produce predictions and what elements underlie the predictions are

significant obstacles in oil and gas exploration. This difficulty arises because many popular machine learning algorithms, including neural networks, are considered “black-box”, meaning they are essentially opaque when explaining their decision-making processes [20].

## 2) INTERPRETABILITY OF MODELS

Owing to the complex and multi-dimensional nature of the data involved in oil and gas exploration, it is difficult to interpret these models. Machine learning algorithms may be trained on a wide range of geophysical data, including seismic surveys, well logs, and production data, all of which can have many characteristics and complicated relationships. Because of the complexity of the data, it may be challenging to interpret the predictions made by machine learning models [49]. It might be difficult to assess model outputs and spot inaccuracies or biases when they are not easily interpretable.

## 3) VISUALIZATION OF HIGH DIMENSIONAL DATA

Visualization of high-dimensional data poses a significant challenge in the oil and gas industry owing to its inherent complexity and issues such as errors, inconsistencies, missing values, and poor data quality. These factors collectively contribute to inaccurate visualizations and limit insights that can be derived from the data. Moreover, when dealing with exceptionally large high-dimensional datasets, computational constraints further exacerbate the difficulties in effectively visualizing the information [127], [128].

## 4) ADDRESSING CHALLENGES

The difficulty in interpreting machine learning models is a significant barrier to penetration in the oil and gas industry, but there are ways to overcome this. This includes feature significance analysis. It examines how each feature in the model’s inputs affects its predictions. Determining which elements are most crucial to the model allows researchers better to comprehend the connections between the data and model predictions.

Using Explainable AI, model interpretation could be improved. Techniques such as SHAP (SHapley Additive exPlanations) values, LIME (Local Interpretable Model-agnostic Explanations), Permutation Importance and Partial Dependence Plot facilitate a deeper understanding of how the model interacts with input characteristics and produces output. By examining these charts, researchers may learn more about how various input variables influence model predictions.

The use of an ensemble model is an alternative approach. A model ensemble aims to provide a more accurate and reliable prediction by merging different machine learning models. Researchers may improve the predictability and clarity of their findings by combining several models with complementary strengths and shortcomings.

Another valuable approach is model visualization, which involves examining the internal mechanisms of a model to gain a deeper understanding of its prediction process. Techniques such as decision tree visualization, activation maximization, and saliency mapping offer insights into hidden connections and patterns within the data that underpin the model’s accuracy. However, special attention is necessary for high-dimensional data to handle the complexities associated with these visualizations effectively. Implementing robust data quality management processes is crucial to ensure accuracy and meaningful insights.

Furthermore, there is a pressing need for advancements in computational capabilities to facilitate the visualization of even larger and more intricate datasets. As the volume and complexity of data continue to grow in the oil and gas industry, it is essential to invest in developing computational resources that can handle the demands of visualizing such vast datasets.

Ultimately, integrating these methods is necessary to overcome the difficulty of interpreting machine learning models in oil and gas exploration by better comprehending the connections between the data and the model’s predictions. Researchers can have more assurance in their estimates and put them to better use in oil and gas exploration if the models are easier to understand. Additionally, by leveraging advancements in visualization, the industry can gain deeper insights into its large and complex data, leading to more informed decision-making processes and improved overall performance.

## C. DOMAIN EXPERTISE

Domain expertise in this context is the familiarity with intricate geology and engineering of oil and gas exploration that comes from years of experience in the field. In addition, there is expertise in machine learning technologies and processes to implement the latest and most effective techniques. Domain knowledge is crucial for ensuring the accuracy and reliability of machine learning models when used in the oil and gas sector.

### 1) EXPERTISE IN OIL AND GAS

Since machine learning models are dependent on input data, the need for domain knowledge arises. Data from seismic surveys, well logs, and production records are all examples of information that may be collected during oil and gas development, all requiring a thorough familiarity with geological and technical fundamentals.

It might be difficult to determine which input characteristics are most important to the machine learning models and whether they correctly represent the underlying geological or engineering processes if one does not have domain knowledge in the field. To guarantee the accuracy of the model’s predictions, domain knowledge is required for calibration and validation.

## 2) EXPERTISE IN MACHINE LEARNING

Expertise in machine learning techniques is essential, as it is in oil and gas. This ensures that an appropriate technique is used at an appropriate time. Machine learning approaches are not one-size-fits-all; each problem and dataset requires a unique solution. Every decision is essential, from model selection to the model optimisation technique. The performance of a machine learning model may depend on the parameters of the model structure that should fit the data. Optimising the parameters in a machine learning model can improve the model's performance. As a result, selecting the best optimisation approaches, including models and evaluation methodologies, is a major challenge that influences the effectiveness and dependability of the industry's machine learning models. Therefore, choosing the optimal optimisation methods, including models and evaluation methodologies, is a major challenge that affects the effectiveness and reliability of the machine learning models of the industry.

## 3) ADDRESSING CHALLENGES

Working collaboratively with domain specialists such as geologists, petrophysicists, and reservoir engineers to train and evaluate machine learning models is one way to overcome this difficulty. This may be done in several ways, including integrating inputs from domain experts throughout the model building process and verifying the models against recognized geological or engineering principles.

Machine learning algorithms that explicitly factor domain expertise are viable options. For instance, some scientists have investigated physics-based machine learning models that leverage well-established scientific principles to enhance model precision and human interpretability.

Overall, the difficulty of domain knowledge in using machine learning in the oil and gas sector highlights the need for data engineers and domain experts to work closely together to ensure that the models are appropriately calibrated, verified and interpreted. If they work together, researchers can create machine learning models that are better suited to the complexity and specialization of the oil and gas industry.

## VIII. CONCLUSION

This study provides an extensive and rigorous examination of machine learning applied within the upstream oil and gas sector, with a particular focus on its pivotal role in the oil and gas exploration domain. Our research endeavours encompass an array of data sources, including meticulously scrutinized research papers, academic theses, and insights shared through conference presentations. A notable concern consistently encountered is the scarcity of data accessible for study in this highly specialized field.

As our investigation underscores, machine learning algorithms have exhibited an extraordinary capability for seismic data processing, accurately classifying facies and lithofacies

and estimating essential petrophysical properties, such as water saturation, permeability, and porosity, across a diverse spectrum of geological formations. The panorama of algorithms employed in these explorations is strikingly diverse, encompassing stalwart techniques like ANN, CNN, SVM, XGBoost, FL, FN, and CM. Notably, the synergy found in hybrid models, which amalgamate multiple algorithms or machine learning models with sophisticated feature selection techniques, consistently offers superior accuracy compared to standalone methodologies.

Despite these promising advancements, several substantial challenges must be addressed for machine learning to reach its full potential in the exploration stage of the oil and gas sector:

- **Data Quality and Availability:** The quality and accessibility of data continue to be a major hurdle. Data in exploration is often limited, unstructured, inconsistent and may contain uncertainties. Solutions must be developed to improve data quality, enhance data sharing among stakeholders, and leverage emerging technologies like remote sensing and real-time data collection.
- **Transparency and Interpretability:** The "black-box" nature of many machine learning models poses challenges in terms of understanding how they arrive at their predictions. Methods for enhancing model transparency and interpretability, such as Explainable AI techniques, must be further explored and integrated into industry practices.
- **Domain Expertise:** Bridging the gap between machine learning experts and domain specialists is essential. Collaboration between data scientists, geologists, petrophysicists, and reservoir engineers is vital to ensure that machine learning models are accurate and aligned with the geological and engineering principles that govern the oil and gas industry.
- **Ethical and Regulatory Considerations:** As with any technology, the use of machine learning in the oil and gas sector must adhere to ethical standards and industry regulations. Addressing data privacy, security, and regulatory compliance is crucial for the responsible application of these powerful tools.

In looking toward the future, several promising directions emerge:

- **Enhancing Robust Machine Learning Frameworks:** The development of robust machine learning frameworks stands as a paramount direction. Oil and gas exploration data is often limited, poor, and compounded by inherent uncertainties. The path forward for machine learning in this domain lies in the creation of adaptive and resilient frameworks. These frameworks should be capable of deriving dependable insights even when faced with the challenges posed by limited, poor, and uncertain data. Such innovation is essential for ensuring the continued efficacy of machine learning in oil and gas exploration.

- **Advanced Visualization:** Innovations in data visualization techniques are critical, especially for handling high-dimensional and complex oil and gas data. Researchers should focus on visual analytics methods that allow for meaningful insights from large and intricate datasets as computational capabilities continue to evolve.
- **Interdisciplinary Collaboration:** Encouraging collaboration between academia, industry, and regulatory bodies can accelerate progress. Joint data sharing, research funding, and standards development efforts can help resolve data quality and access issues.
- **Regulatory Compliance Tools:** The development of tools and frameworks that assist in navigating the complex regulatory landscape of the oil and gas industry is essential. These tools should facilitate compliance while ensuring data security and privacy.
- **Computational Capabilities:** Continued investment in computational resources is vital to handle the increasing volume of data and the computational demands of machine learning algorithms. This includes exploring cloud computing, distributed computing, and high-performance computing solutions.

The review contributes significantly to understanding the unique challenges in applying machine learning to the exploration stage in the oil and gas industry, such as uncertainties in exploration parameters, scale discrepancies, and complexities in handling temporal and spatial data. Notably, the review goes beyond identification; it offers potential solutions, identifies practices contributing to achieving optimal accuracy, and outlines future research directions, providing a nuanced understanding of the field's dynamics. This comprehensive analysis provides a roadmap for overcoming challenges and enriching the knowledge base for researchers and industry stakeholders.

## REFERENCES

- [1] M. Mohammadpoor and F. Torabi, "Big data analytics in oil and gas industry: An emerging trend," *Petroleum*, vol. 6, no. 4, pp. 321–328, Dec. 2020.
- [2] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, "Application of machine learning and artificial intelligence in oil and gas industry," *Petroleum Res.*, vol. 6, no. 4, pp. 379–391, Dec. 2021.
- [3] 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, Nov. 2019, Art. no. 100191.
- [4] L. Kuang, H. Liu, Y. Ren, K. Luo, M. Shi, J. Su, and X. Li, "Application and development trend of artificial intelligence in petroleum exploration and development," *Petroleum Explor. Develop.*, vol. 48, no. 1, pp. 1–14, Feb. 2021.
- [5] R. K. Pandey, A. K. Dahiya, and A. Mandal, "Identifying applications of machine learning and data analytics based approaches for optimization of upstream petroleum operations," *Energy Technol.*, vol. 9, no. 1, Jan. 2021, Art. no. 2000749.
- [6] R. K. Pandey, H. Kakati, and A. Mandal, "Thermodynamic modeling of equilibrium conditions of CH<sub>4</sub>/CO<sub>2</sub>/N<sub>2</sub> clathrate hydrate in presence of aqueous solution of sodium chloride inhibitor," *Petroleum Sci. Technol.*, vol. 35, no. 10, pp. 947–954, May 2017.
- [7] A. Guitton, H. Wang, and W. Trainor-Guitton, "Statistical imaging of faults in 3D seismic volumes using a machine learning approach," in *Proc. SEG Tech. Program Expanded Abstr.* Houston, TX, USA: Society of Exploration Geophysicists, Aug. 2017, pp. 2045–2049.
- [8] W. Xiong, X. Ji, Y. Ma, Y. Wang, N. M. Al BinHassan, M. N. Ali, and Y. Luo, "Seismic fault detection with convolutional neural network," *Geophysics*, vol. 83, no. 5, pp. O97–O103, Sep. 2018.
- [9] H. Di and D. Gao, "Gray-level transformation and Canny edge detection for 3D seismic discontinuity enhancement," *Comput. Geosci.*, vol. 72, pp. 192–200, Nov. 2014.
- [10] H. Di and D. Gao, "Improved estimates of seismic curvature and flexure based on 3D surface rotation in the presence of structure dip," *Geophysics*, vol. 81, no. 2, pp. IM13–IM23, Mar. 2016.
- [11] L. Yang and S. Z. Sun, "Seismic horizon tracking using a deep convolutional neural network," *J. Petroleum Sci. Eng.*, vol. 187, Apr. 2020, Art. no. 106709.
- [12] S. Diersen, E.-J. Lee, D. Spears, P. Chen, and L. Wang, "Classification of seismic windows using artificial neural networks," *Proc. Comput. Sci.*, vol. 4, pp. 1572–1581, Jan. 2011.
- [13] M. Rastegarnia, A. Sanati, and D. Javani, "A comparative study of 3D FZI and electrofacies modeling using seismic attribute analysis and neural network technique: A case study of Cheshmeh-Khosh oil field in Iran," *Petroleum*, vol. 2, no. 3, pp. 225–235, Sep. 2016.
- [14] K. La Marca, H. Bedle, L. Stright, R. P. de Lima, and K. J. Marfurt, "Quantifying uncertainty in unsupervised machine learning methods for seismic facies using outcrop-derived 3D models and synthetic seismic data," in *Proc. SEG Int. Expo. Annu. Meeting*, 2022, pp. 1312–1316.
- [15] K.-R. Qian, Z.-L. He, X.-W. Liu, and Y.-Q. Chen, "Intelligent prediction and integral analysis of shale oil and gas sweet spots," *Petroleum Sci.*, vol. 15, no. 4, pp. 744–755, Nov. 2018.
- [16] Y. Xie, C. Zhu, W. Zhou, Z. Li, X. Liu, and M. Tu, "Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances," *J. Petroleum Sci. Eng.*, vol. 160, pp. 182–193, Jan. 2018.
- [17] P. Avseth and T. Mukerji, "Seismic lithofacies classification from well logs using statistical rock physics," *Petrophysics-SPWLA J. Formation Eval. Reservoir Description*, vol. 43, no. 2, 2002.
- [18] P. Delfiner, O. Peyret, and O. Serra, "Automatic determination of lithology from well logs," *SPE Formation Eval.*, vol. 2, no. 3, pp. 303–310, Sep. 1987.
- [19] A. A. Silva, I. A. L. Neto, R. M. Misságia, M. A. Ceia, A. G. Carrasquilla, and N. L. Archilha, "Artificial neural networks to support petrographic classification of carbonate-siliciclastic rocks using well logs and textural information," *J. Appl. Geophys.*, vol. 117, pp. 118–125, Jun. 2015.
- [20] Z. Tariq, M. S. Aljawad, A. Hasan, M. Murtaza, E. Mohammed, A. El-Husseiny, S. A. Alarifi, M. Mahmoud, and A. Abdulaheem, "A systematic review of data science and machine learning applications to the oil and gas industry," *J. Petroleum Explor. Prod. Technol.*, vol. 11, no. 12, pp. 4339–4374, 2021.
- [21] J. L. Baldwin, R. M. Bateman, and C. L. Wheatley, "Application of a neural network to the problem of mineral identification from well logs," *Log Analyst*, vol. 31, no. 5, 1990.
- [22] L. Qi and T. R. Carr, "Neural network prediction of carbonate lithofacies from well logs, big bow and sand arroyo creek fields, southwest Kansas," *Comput. Geosci.*, vol. 32, no. 7, pp. 947–964, Aug. 2006.
- [23] M. A. Sebtoosheikh and A. Salehi, "Lithology prediction by support vector classifiers using inverted seismic attributes data and petrophysical logs as a new approach and investigation of training data set size effect on its performance in a heterogeneous carbonate reservoir," *J. Petroleum Sci. Eng.*, vol. 134, pp. 143–149, Oct. 2015.
- [24] P. Dell'Aversana, "Comparison of different machine learning algorithms for lithofacies classification from well logs," *Bollettino di Geofisica Teorica ed Applicata*, vol. 60, no. 1, pp. 69–80, 2019.
- [25] E. Lutins, "Ensemble methods in machine learning: What are they and why use them," Graduate Current Data Scientist at Pinpoint, Univ. Maryland, College Park, MD, USA, Tech. Rep., 2017.
- [26] S. Tewari and U. D. Dwivedi, "A comparative study of heterogeneous ensemble methods for the identification of geological lithofacies," *J. Petroleum Explor. Prod. Technol.*, vol. 10, no. 5, pp. 1849–1868, Jun. 2020.
- [27] S. Bhattacharya and S. Mishra, "Applications of machine learning for facies and fracture prediction using Bayesian network theory and random forest: Case studies from the appalachian basin, USA," *J. Petroleum Sci. Eng.*, vol. 170, pp. 1005–1017, Nov. 2018.
- [28] M. Hou, Y. Xiao, Z. Lei, Z. Yang, Y. Lou, and Y. Liu, "Machine learning algorithms for lithofacies classification of the Gulong Shale from the Songliao Basin, China," *Energies*, vol. 16, no. 6, p. 2581, Mar. 2023.

- [29] W. J. Al-Mudhafar, M. A. Abbas, and D. A. Wood, "Performance evaluation of boosting machine learning algorithms for lithofacies classification in heterogeneous carbonate reservoirs," *Mar. Petroleum Geol.*, vol. 145, Nov. 2022, Art. no. 105886.
- [30] J. Kim, "Lithofacies classification integrating conventional approaches and machine learning technique," *J. Natural Gas Sci. Eng.*, vol. 100, Apr. 2022, Art. no. 104500.
- [31] D. Datta, G. Singh, S. K. Singh, M. Jenamani, and A. Routray, "Application of multivariate change detection in automated lithofacies classification from well-log data in a nonstationary subsurface," *J. Appl. Geophys.*, vol. 215, Aug. 2023, Art. no. 105094.
- [32] L. Lake, *Reservoir Characterization*. Amsterdam, The Netherlands: Elsevier, 2012.
- [33] R. C. Selley and S. A. Sonnenberg, "The reservoir," in *Elements of Petroleum Geology*, R. C. Selley and S. A. Sonnenberg, Eds., 3rd ed. Boston, MA, USA: Academic, pp. 255–320, 2015, ch. 6.
- [34] A. Adeniran, M. Elshafei, and G. Hamada, "Functional network softsensor for formation porosity and water saturation in oil wells," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, May 2009, pp. 1138–1143.
- [35] M. Kennedy, *Practical Petrophysics*. Amsterdam, The Netherlands: Elsevier, 2015.
- [36] C. Xu, S. Misra, P. Srinivasan, and S. Ma, "When petrophysics meets big data: What can machine do?" in *Proc. SPE Middle East Oil Gas Show Conf.*, 2019, Paper SPE-195068-MS.
- [37] *Robust Estimation of Reservoir Properties—How to Turn a Week Into Minutes With Machine Learning*, PGS.
- [38] Z. Huang, J. Shimeld, M. Williamson, and J. Katsube, "Permeability prediction with artificial neural network modeling in the venture gas field, offshore eastern Canada," *Geophysics*, vol. 61, no. 2, pp. 422–436, Mar. 1996.
- [39] H. B. Helle, A. Bhatt, and B. Ursin, "Porosity and permeability prediction from wireline logs using artificial neural networks: A north sea case study," *Geophys. Prospecting*, vol. 49, no. 4, pp. 431–444, Jul. 2001.
- [40] S. Singh, "Permeability prediction using artificial neural network (ANN): A case study of Uinta basin," in *Proc. SPE Annu. Tech. Conf. Exhib.*, 2005, Paper SPE-99286.
- [41] M. Abdideh, "Estimation of permeability using artificial neural networks and regression analysis in an Iran oil field," *Int. J. Phys. Sci.*, vol. 7, no. 34, pp. 5308–5313, Sep. 2012.
- [42] J. Ben-Awuah and E. Padmanabhan, "An enhanced approach to predict permeability in reservoir sandstones using artificial neural networks (ANN)," *Arabian J. Geosci.*, vol. 10, no. 7, pp. 1–15, Apr. 2017.
- [43] S. Elkhatny, M. Mahmoud, Z. Tariq, and A. Abdurraheem, "New insights into the prediction of heterogeneous carbonate reservoir permeability from well logs using artificial intelligence network," *Neural Comput. Appl.*, vol. 30, no. 9, pp. 2673–2683, Nov. 2018.
- [44] B. Basbug and Z. T. Karpyn, "Estimation of permeability from porosity, specific surface area, and irreducible water saturation using an artificial neural network," in *Proc. Latin Amer. Caribbean Petroleum Eng. Conf.*, 2007, Paper SPE-107909.
- [45] R. Irani and R. Nasimi, "Evolving neural network using real coded genetic algorithm for permeability estimation of the reservoir," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9862–9866, Aug. 2011.
- [46] P. Bagheripour, "Committee neural network model for rock permeability prediction," *J. Appl. Geophys.*, vol. 104, pp. 142–148, May 2014.
- [47] M. Matinkia, R. Hashami, M. Mehrad, M. R. Hajsaeedi, and A. Velayati, "Prediction of permeability from well logs using a new hybrid machine learning algorithm," *Petroleum*, vol. 9, no. 1, pp. 108–123, Mar. 2023.
- [48] X. Zhao, X. Chen, Q. Huang, Z. Lan, X. Wang, and G. Yao, "Logging-data-driven permeability prediction in low-permeable sandstones based on machine learning with pattern visualization: A case study in Wenchang A Sag, Pearl River Mouth Basin," *J. Petroleum Sci. Eng.*, vol. 214, Jul. 2022, Art. no. 110517.
- [49] J.-J. Liu and J.-C. Liu, "Permeability predictions for tight sandstone reservoir using explainable machine learning and particle swarm optimization," *Geofluids*, vol. 2022, pp. 1–15, Jan. 2022.
- [50] P. Tahmasebi and A. Hezarkhani, "A fast and independent architecture of artificial neural network for permeability prediction," *J. Petroleum Sci. Eng.*, vols. 86–87, pp. 118–126, May 2012.
- [51] M. Jamialahmadi and F. G. Javadpour, "Relationship of permeability, porosity and depth using an artificial neural network," *J. Petroleum Sci. Eng.*, vol. 26, nos. 1–4, pp. 235–239, May 2000.
- [52] H. Kaydani, A. Mohebbi, and A. Baghaie, "Permeability prediction based on reservoir zonation by a hybrid neural genetic algorithm in one of the Iranian heterogeneous oil reservoirs," *J. Petroleum Sci. Eng.*, vol. 78, no. 2, pp. 497–504, Aug. 2011.
- [53] T. Aifa, R. Baouche, and K. Baddari, "Neuro-fuzzy system to predict permeability and porosity from well log data: A case study of Hassi R'Mel gas field, Algeria," *J. Petroleum Sci. Eng.*, vol. 123, pp. 217–229, Nov. 2014.
- [54] B. S. Saljooghi and A. Hezarkhani, "A new approach to improve permeability prediction of petroleum reservoirs using neural network adaptive wavelet (wavenet)," *J. Petroleum Sci. Eng.*, vol. 133, pp. 851–861, Sep. 2015.
- [55] S. Baziar, M. Tadayoni, M. Nabi-Bidhendi, and M. Khalili, "Prediction of permeability in a tight gas reservoir by using three soft computing approaches: A comparative study," *J. Natural Gas Sci. Eng.*, vol. 21, pp. 718–724, Nov. 2014.
- [56] M. Z. Kamali, S. Davoodi, H. Ghorbani, D. A. Wood, N. Mohamadian, S. Lajmorak, V. S. Rukavishnikov, F. Taherizade, and S. S. Band, "Permeability prediction of heterogeneous carbonate gas condensate reservoirs applying group method of data handling," *Mar. Petroleum Geol.*, vol. 139, May 2022, Art. no. 105597.
- [57] G. Hamada and M. Elshafei, "Neural network prediction of porosity and permeability of heterogeneous gas sand reservoirs," in *Proc. SPE Saudi Arabia Section Tech. Symp.*, 2009, Paper SPE-126042.
- [58] E. A. El-Sebakhy, O. Asparouhov, A.-A. Abdurraheem, A.-A. Al-Majed, D. Wu, K. Latinski, and I. Raharja, "Functional networks as a new data mining predictive paradigm to predict permeability in a carbonate reservoir," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10359–10375, Sep. 2012.
- [59] S. O. Olatunji, A. Selamat, and A. A. A. Raheem, "Extreme learning machines based model for predicting permeability of carbonate reservoir," *Int. J. Digit. Content Technol. Appl.*, vol. 7, no. 1, p. 450, 2013.
- [60] R. Gholami, A. Moradzadeh, S. Maleki, S. Amiri, and J. Hanachi, "Applications of artificial intelligence methods in prediction of permeability in hydrocarbon reservoirs," *J. Petroleum Sci. Eng.*, vol. 122, pp. 643–656, Oct. 2014.
- [61] A. Abdurraheem, E. Sabakhy, M. Ahmed, A. Vantala, P. D. Raharja, and G. Korvin, "Estimation of permeability from wireline logs in a middle eastern carbonate reservoir using fuzzy logic," in *Proc. SPE Middle East Oil Gas Show Conf.*, 2007, Paper SPE-105350.
- [62] X. Wang, S. Yang, Y. Wang, Y. Zhao, and B. Ma, "Improved permeability prediction based on the feature engineering of petrophysics and fuzzy logic analysis in low porosity-permeability reservoir," *J. Petroleum Explor. Prod. Technol.*, vol. 9, no. 2, pp. 869–887, Jun. 2019.
- [63] G. Zhang, Z. Wang, H. Li, Y. Sun, Q. Zhang, and W. Chen, "Permeability prediction of isolated channel sands using machine learning," *J. Appl. Geophys.*, vol. 159, pp. 605–615, Dec. 2018.
- [64] A. Sheykhasab, A. A. Mohseni, A. B. Bahari, E. Naruei, S. Davoodi, A. Aghaz, and M. Mehrad, "Prediction of permeability of highly heterogeneous hydrocarbon reservoir from conventional petrophysical logs using optimized data-driven algorithms," *J. Petroleum Explor. Prod. Technol.*, vol. 13, no. 2, pp. 661–689, Feb. 2023.
- [65] F. A. Anifowose, A. Abdurraheem, A. Al-Shuhail, and D. P. Schmitt, "Improved permeability prediction from seismic and log data using artificial intelligence techniques," in *Proc. SPE Middle East Oil Gas Show Conf.*, 2013, Paper SPE-164465.
- [66] A. Bhatt, "Reservoir properties from well logs using neural networks," Ph.D. thesis, Dept. Petroleum Eng. Appl. Geophys., Citeseer, Norwegian Univ. Sci. Technol., Trondheim, Norway, 2002.
- [67] C.-H. Chen and Z.-S. Lin, "A committee machine with empirical formulas for permeability prediction," *Comput. Geosci.*, vol. 32, no. 4, pp. 485–496, May 2006.
- [68] M. R. J. Wyllie and W. D. Rose, "Some theoretical considerations related to the quantitative evaluation of the physical characteristics of reservoir rock from electrical log data," *J. Petroleum Technol.*, vol. 2, no. 04, pp. 105–118, Apr. 1950.
- [69] G. R. Coates and J. Dumanoir, "A new approach to improved log-derived permeability," in *Proc. SPWLA 14th Annu. Logging Symp.*, 1973, Paper SPWLA-1973.
- [70] *Principles/Applications*, L Interpretation, Schlumberger Educ. Services, Houston, TX, USA, 1989.

- [71] R. Sadeghi, A. Kadkhodaie, B. Rafiei, M. Yosefpour, and S. Khodabakhsh, "A committee machine approach for predicting permeability from well log data: A case study from a heterogeneous carbonate reservoir, Balal oil field, Persian Gulf," *Geopersia*, vol. 1, no. 2, pp. 1–10, 2011.
- [72] T. Helmy, S. M. Rahman, M. I. Hossain, and A. Abdelraheem, "Non-linear heterogeneous ensemble model for permeability prediction of oil reservoirs," *Arabian J. Sci. Eng.*, vol. 38, no. 6, pp. 1379–1395, Jun. 2013.
- [73] F. A. Anifowose, J. Labadin, and A. Abdullaheem, "Non-linear feature selection-based hybrid computational intelligence models for improved natural gas reservoir characterization," *J. Natural Gas Sci. Eng.*, vol. 21, pp. 397–410, Nov. 2014.
- [74] M. Masroor, M. E. Niri, and M. H. Sharifinasab, "A multiple-input deep residual convolutional neural network for reservoir permeability prediction," *Geoenergy Sci. Eng.*, vol. 222, Mar. 2023, Art. no. 211420.
- [75] F. Anifowose, J. Labadin, and A. Abdullaheem, "Ensemble learning model for petroleum reservoir characterization: A case of feed-forward back-propagation neural networks," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Gold Coast, QLD, Australia: Springer, 2013, pp. 71–82.
- [76] F. Anifowose, J. Labadin, and A. Abdullaheem, "Ensemble model of artificial neural networks with randomized number of hidden neurons," in *Proc. 8th Int. Conf. Inf. Technol. Asia (CITA)*, Jul. 2013, pp. 1–5.
- [77] F. Anifowose, J. Labadin, and A. Abdullaheem, "Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines," *Appl. Soft Comput.*, vol. 26, pp. 483–496, Jan. 2015.
- [78] F. A. Anifowose, J. Labadin, and A. Abdullaheem, "Ensemble model of non-linear feature selection-based extreme learning machine for improved natural gas reservoir characterization," *J. Natural Gas Sci. Eng.*, vol. 26, pp. 1561–1572, Sep. 2015.
- [79] D. A. Otchere, T. O. A. Ganat, R. Gholami, and M. Lawal, "A novel custom ensemble learning model for an improved reservoir permeability and water saturation prediction," *J. Natural Gas Sci. Eng.*, vol. 91, Jul. 2021, Art. no. 103962.
- [80] J.-S. Lim and J. Kim, "Reservoir porosity and permeability estimation from well logs using fuzzy logic and neural networks," in *Proc. SPE Asia Pacific Oil Gas Conf. Exhib.*, 2004, Paper SPE-88476.
- [81] H. Zargari, S. Poordad, and R. Kharrat, "Porosity and permeability prediction based on computational intelligences as artificial neural networks (ANNs) and adaptive neuro-fuzzy inference systems (ANFIS) in southern carbonate reservoir of Iran," *Petroleum Sci. Technol.*, vol. 31, no. 10, pp. 1066–1077, May 2013.
- [82] A. A. Konaté, H. Pan, N. Khan, and J. H. Yang, "Generalized regression and feed-forward back propagation neural networks in modelling porosity from geophysical well logs," *J. Petroleum Explor. Prod. Technol.*, vol. 5, no. 2, pp. 157–166, Jun. 2015.
- [83] S. Elkatatny, Z. Tariq, M. Mahmoud, and A. Abdullaheem, "New insights into porosity determination using artificial intelligence techniques for carbonate reservoirs," *Petroleum*, vol. 4, no. 4, pp. 408–418, Dec. 2018.
- [84] M. Nourani, N. Alali, S. Samadianfard, S. S. Band, K.-W. Chau, and C.-M. Shu, "Comparison of machine learning techniques for predicting porosity of chalk," *J. Petroleum Sci. Eng.*, vol. 209, Feb. 2022, Art. no. 109853.
- [85] Z. Zhang, Y. Wang, and P. Wang, "On a deep learning method of estimating reservoir porosity," *Math. Problems Eng.*, vol. 2021, pp. 1–13, Feb. 2021.
- [86] M. A. Ahmadi and Z. Chen, "Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petrophysical logs," *Petroleum*, vol. 5, no. 3, pp. 271–284, Sep. 2019.
- [87] Y. Sun, J. Zhang, Z. Yu, Z. Liu, and P. Yin, "WOA (whale optimization algorithm) optimizes Elman neural network model to predict porosity value in well logging curve," *Energies*, vol. 15, no. 12, p. 4456, 2022.
- [88] J. Wang and J. Cao, "Deep learning reservoir porosity prediction using integrated neural network," *Arabian J. Sci. Eng.*, vol. 47, no. 9, pp. 11313–11327, Sep. 2022.
- [89] A. F. Al-Anazi and I. D. Gates, "Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study," *Comput. Geosci.*, vol. 36, no. 12, pp. 1494–1503, Dec. 2010.
- [90] M.-A. Ahmadi, M. R. Ahmadi, S. M. Hosseini, and M. Ebadi, "Connectionist model predicts the porosity and permeability of petroleum reservoirs by means of petro-physical logs: Application of artificial intelligence," *J. Petroleum Sci. Eng.*, vol. 123, pp. 183–200, Nov. 2014.
- [91] Z. Zhong and T. R. Carr, "Application of a new hybrid particle swarm optimization-mixed kernels function-based support vector machine model for reservoir porosity prediction: A case study in Jacksonburg-Stringtown oil field, West Virginia, USA," *Interpretation*, vol. 7, no. 1, pp. T97–T112, Feb. 2019.
- [92] P. Ø. Andersen, M. Skjeldal, and C. Augustsson, "Machine learning based prediction of porosity and water saturation from varg field reservoir well logs," in *Proc. SPE Europec-Eur. Energy Conf. Featured 83rd EAGE Annu. Conf. Exhib.*, 2022.
- [93] Z. Tariq, M. Gudala, B. Yan, S. Sun, and M. Mahmoud, "A fast method to infer nuclear magnetic resonance based effective porosity in carbonate rocks using machine learning techniques," *Geoenergy Sci. Eng.*, vol. 222, Mar. 2023, Art. no. 211333.
- [94] M. F. Haqqi, S. Saroji, and S. Prakoso, "An implementation of XGBoost algorithm to estimate effective porosity on well log data," *J. Phys., Conf. Ser.*, vol. 2498, no. 1, May 2023, Art. no. 012011.
- [95] S. Pan, Z. Zheng, Z. Guo, and H. Luo, "An optimized XGBoost method for predicting reservoir porosity using petrophysical logs," *J. Petroleum Sci. Eng.*, vol. 208, Jan. 2022, Art. no. 109520.
- [96] M. I. Miah, S. Zendejboudi, and S. Ahmed, "Log data-driven model and feature ranking for water saturation prediction using machine learning approach," *J. Petroleum Sci. Eng.*, vol. 194, Nov. 2020, Art. no. 107291.
- [97] G. E. Archie, "The electrical resistivity log as an aid in determining some reservoir characteristics," *Trans. AIME*, vol. 146, no. 1, pp. 54–62, Dec. 1942.
- [98] P. Simandoux, "Dielectric measurements on porous media, application to the measurements of water saturation: Study of behavior of argillaceous formations," *Revue de L'institut Fr. du Petrole*, vol. 18, pp. 193–215, 1963.
- [99] A. Poupon, M. E. Loy, and M. P. Tixier, "A contribution to electrical log interpretation in shaly sands," *J. Petroleum Technol.*, vol. 6, no. 6, pp. 27–34, Jun. 1954.
- [100] L. de Witte, "Relations between resistivities and fluid contents of porous rocks," *Oil Gas J.*, vol. 49, no. 16, pp. 120–134, 1950.
- [101] M. H. Waxman and L. J. M. Smits, "Electrical conductivities in oil-bearing shaly sands," *Soc. Petroleum Eng. J.*, vol. 8, no. 2, pp. 107–122, Jun. 1968.
- [102] W. H. Fertl and G. W. Hammack, "A comparative look at water saturation computations in shaly pay sands," in *Proc. SPWLA 12th Annu. Logging Symp.*, 1971, Paper SPWLA-1971.
- [103] H. B. Helle and A. Bhatt, "Fluid saturation from well logs using committee neural networks," *Petroleum Geosci.*, vol. 8, no. 2, pp. 109–118, Jun. 2002.
- [104] E. E.-M. Shokir, "Prediction of the hydrocarbon saturation in low resistivity formation via artificial neural network," in *Proc. SPE Asia Pacific Conf. Integr. Modelling Asset Manag.*, 2004, Paper SPE-87001.
- [105] K. Kamalyar, "Using artificial neural network for predicting water saturation in an Iranian oil reservoir," in *Proc. 10th EAGE Int. Conf. Geoinform.-Theor. Appl. Aspects*. Amsterdam, The Netherlands: European Association of Geoscientists & Engineers, 2011, p. cp-240.
- [106] N. Al-Bulushi, P. R. King, M. J. Blunt, and M. Kraaijveld, "Development of artificial neural network models for predicting water saturation and fluid distribution," *J. Petroleum Sci. Eng.*, vol. 68, nos. 3–4, pp. 197–208, Oct. 2009.
- [107] M. Mardi, H. Nurozi, and S. Edalatkhah, "A water saturation prediction using artificial neural networks and an investigation on cementation factors and saturation exponent variations in an Iranian oil well," *Petroleum Sci. Technol.*, vol. 30, no. 4, pp. 425–434, Feb. 2012.
- [108] A. N. Okon, S. E. Adewole, and E. M. Uguma, "Artificial neural network model for reservoir petrophysical properties: Porosity, permeability and water saturation prediction," *Model. Earth Syst. Environ.*, vol. 7, no. 4, pp. 2373–2390, Nov. 2021.
- [109] N. I. Al-Bulushi, P. R. King, M. J. Blunt, and M. Kraaijveld, "Artificial neural networks workflow and its application in the petroleum industry," *Neural Comput. Appl.*, vol. 21, no. 3, pp. 409–421, Apr. 2012.
- [110] C. Y. Nyein, G. M. Hamada, and A. Elsakka, "Artificial neural network (ANN) prediction of porosity and water saturation of shaly sandstone reservoirs," in *Proc. AAPG Asia Pacific Region, 4th AAPG/EAGE/MGS Myanmar Oil Gas Conf., Myanmar, Global Oil Gas Hotspot, Unleashing Petroleum Syst. Potential*, 2018.
- [111] S. A. Jafari Kenari and S. Mashohor, "Robust committee machine for water saturation prediction," *J. Petroleum Sci. Eng.*, vol. 104, pp. 1–10, Apr. 2013.

- [112] A. F. Ibrahim, S. Elkhatatny, and M. Al Ramadan, "Prediction of water saturation in tight gas sandstone formation using artificial intelligence," *ACS Omega*, vol. 7, no. 1, pp. 215–222, Jan. 2022.
- [113] M. R. Khan, Z. Tariq, and A. Abdurraheem, "Machine learning derived correlation to determine water saturation in complex lithologies," in *Proc. SPE Kingdom Saudi Arabia Annu. Tech. Symp. Exhib.*, 2018, Paper SPE-192307.
- [114] B. Bageri, F. Anifowose, and A. Abdurraheem, "Artificial intelligence based estimation of water saturation using electrical measurements data in a carbonate reservoir," in *Proc. SPE Middle East Oil Gas Show Conf.*, 2015, Paper SPE-172564.
- [115] M. Amiri, J. Ghiasi-Freez, B. Golkar, and A. Hatampour, "Improving water saturation estimation in a tight shaly sandstone reservoir using artificial neural network optimized by imperialist competitive algorithm—A case study," *J. Petroleum Sci. Eng.*, vol. 127, pp. 347–358, Mar. 2015.
- [116] H. H. Gholanlo, M. Amirpour, and S. Ahmadi, "Estimation of water saturation by using radial based function artificial neural network in carbonate reservoir: A case study in Sarvak formation," *Petroleum*, vol. 2, no. 2, pp. 166–170, Jun. 2016.
- [117] Z. Tariq, M. Mahmoud, and A. Abdurraheem, "An intelligent data-driven model for dean–stark water saturation prediction in carbonate rocks," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 11919–11935, Aug. 2020.
- [118] A. Mollajan, H. Memarian, and M. Jalali, "Prediction of reservoir water saturation using support vector regression in an Iranian carbonate reservoir," in *Proc. 47th U.S. Rock Mech./Geomechanics Symp.*, 2013, Paper ARMA-2013.
- [119] B. Zhao, H. Zhou, X. Li, and D. Han, "Water saturation estimation using support vector machine," in *Proc. SEG Tech. Program Expanded Abstr.* Houston, TX, USA: Society of Exploration Geophysicists, 2006, pp. 1693–1697.
- [120] S. Baziar, H. B. Shahripour, M. Tadayoni, and M. Nabi-Bidhendi, "Prediction of water saturation in a tight gas sandstone reservoir by using four intelligent methods: A comparative study," *Neural Comput. Appl.*, vol. 30, no. 4, pp. 1171–1185, Aug. 2018.
- [121] F. Hadavimoghaddam, M. Ostadhassan, M. A. Sadri, T. Bondarenko, I. Chebyshev, and A. Semnani, "Prediction of water saturation from well log data by machine learning algorithms: Boosting and super learner," *J. Mar. Sci. Eng.*, vol. 9, no. 6, p. 666, Jun. 2021.
- [122] R. Copus, R. Hubert, and H. Laqueur, "Credible prediction: Big data, machine learning and the credibility revolution," *Mach. Learn. Credibility Revolution*, pp. 22–57, Apr. 2018.
- [123] Y. Hajizadeh, "Machine learning in oil and gas; a SWOT analysis approach," *J. Petroleum Sci. Eng.*, vol. 176, pp. 661–663, May 2019.
- [124] D. Koroteev and Z. Tekic, "Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future," *Energy AI*, vol. 3, Mar. 2021, Art. no. 100041.
- [125] A. Mathrani, T. Susnjak, G. Ramaswami, and A. Barczak, "Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics," *Comput. Educ. Open*, vol. 2, Dec. 2021, Art. no. 100060.
- [126] M. Frye, J. Mohren, and R. H. Schmitt, "Benchmarking of data preprocessing methods for machine learning-applications in production," *Proc. CIRP*, vol. 104, pp. 50–55, Jan. 2021.
- [127] A. E. Mostafa, "PetroVis & FractVis: Interactive visual exploration of high-dimensional oil and gas data," Ph.D. thesis, Dept. Comput. Sci., Univ. Calgary, Calgary, AB, Canada, 2013.
- [128] A. Mazher, "Visualization framework for high-dimensional spatio-temporal hydrological gridded datasets using machine-learning techniques," *Water*, vol. 12, no. 2, p. 590, Feb. 2020.



**AHMAD LAWAL** received the B.Sc. degree in mathematics from Usmanu Danfodiyo University, Sokoto, Nigeria, in 2018, and the M.Sc. degree in software engineering from De Montfort University, Leicester, U.K., in 2020, where he is currently pursuing the Ph.D. degree, with a focus on enhancing reservoir characterization using machine learning.

He is holding a significant role as a part-time Lecturer in computing. Simultaneously, he contributes as a Research Fellow for the ADRELO project, funded by EPSRC/UKRI. In this capacity, his focus centers on leveraging time series machine learning algorithms for advanced applications. His research interests include machine learning and data analytics.



**YINGJIE YANG** received the B.Sc. (Hons.), M.Sc., and Ph.D. degrees in engineering from Northeastern University, Shenyang, China, in 1987, 1990, and 1994, respectively, and the Ph.D. degree in computer science from Loughborough University, Loughborough, U.K., in 2008.

He is currently a Professor of computational intelligence with De Montfort University, Leicester, U.K. His research has been supported by the Royal Society, EU FP7, European Space Agency, Leverhulme Trust, and De Montfort University. He has published over 200 papers on grey systems, fuzzy sets, rough sets, neural networks, and their applications to civil engineering, transportation, environmental engineering, and management science.

Prof. Yang has served as a PC member for over 100 international conferences. He is the Co-Chair of the IEEE Technical Committee on Grey Systems in the IEEE Systems, Man and Cybernetics Society. He is the Executive President of the International Association on Grey Systems and Uncertainty Analysis. He is serving as an Associate Editor for six academic journals, including IEEE TRANSACTIONS ON CYBERNETICS.



**HONGMEI HE** received the Ph.D. degree in computer science from Loughborough University, U.K., in 2006.

After the Ph.D. degree, she worked at various universities, such as the University of Bristol, Ulster University, the University of Kent, Cranfield University, and De Montfort University. She is currently a Professor of future robotics, engineering and transport systems with the School of Science, Engineering, and Environment, University of Salford, Manchester, U.K. Before coming to the U.K., she was a Senior Embedded Systems Engineer with Motorola Design House, China. Her research is primarily funded by EPSRC, Innovation U.K., the Ministry of Defence, and industry. Her research interests include AI for trustworthy robotics and autonomous systems (TRAS) in terms of safety, security, human–robot interaction, system health, and ethics.

Prof. She is also a Working Group Member of the IEEE P1912 Standard for Privacy and Security Framework for Consumer Wireless Devices. She is the Chair of the IEEE U.K. and Ireland RAS Chapter and the Chair of the AI and Edge Computing for TRAS Task Force in the Adaptive Dynamic Programming and Reinforcement Learning Technical Committee (ADPRLTC) of the IEEE Computational Intelligence Society.



**NATHANAEL L. BAISA** received the B.Sc. degree (Hons.) in electrical engineering from Mekelle University, Ethiopia, in 2008, and the European Erasmus Mundus M.Sc. degree (Hons.) in computer vision and robotics (VIBOT) from Burgundy University, France, Girona University, Spain, and Heriot-Watt University, U.K., in 2013, and the Ph.D. degree in electrical engineering from Heriot-Watt University, in 2018.

He is currently a Lecturer (Assistant Professor) in artificial intelligence (AI) with De Montfort University. Before joining De Montfort University, he was a Senior Research Associate with Lancaster University, a Research Scientist with AnyVision, and a Research Fellow with the University of Lincoln, where he participated as a main researcher in many research projects funded by ERC (under EU's Horizon 2020), EPSRC, and Innovate U.K. His research interests include computer vision and machine learning.

...