SEMARAK ILMU PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

An In-Depth Review of Predictive Methods for Oil and Gas Applications

Marina Yusoff[1,2,4,*], Mohamad Taufik Mohd Sallehud-din[3], Putri Azmira R. Azmi[1], Nor Hapiza Mohd Ariffin[4], Calvin Karunakumar[3]

[1] Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Kompleks Al-Khawarizmi, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
[2] College of Computing, Informatic and M, Kompleks Al-Khawarizmi, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
[3] PETRONAS Research Sdn Bhd, Kawasan Institusi Bangi, 43000 Bandar Baru Bangi, Selangor, Malaysia
[4] MIS Department, Faculty of Business, Sohar University, Sohar 311, Oman

ARTICLE INFO

Article history:
Received 8 December 2023
Received in revised form 29 July 2024
Accepted 1 August 2024
Available online 20 August 2024

Keywords:
Deep learning; machine learning; predictive analytics; oil and gas; SLR

ABSTRACT

The oil and gas industry aims to optimize production, reduce costs, and increase efficiency. Predictive models have gained popularity as potential solutions in recent years. In predictive models, machine learning algorithms analyse oil and gas operations data and forecast future performance. This paper examines the current state of predictive modelling in the oil and gas industry with the objective to systematically review and analyse current research on the predictive modelling in the oil and gas industry. The paper begins by highlighting the sub-fields and datasets in the oil and gas industry that used recent machine learning methods for predictive modelling. Additionally, literature from the Scopus and Web of Science indexes was reviewed. This study assessed recent approaches for oil and gas industry in predictive applications for papers published up until December 2022. The findings identify several advantages and disadvantages that can be used as guidelines to effectively implement predictive modelling in the oil and gas industry. It includes challenges on the requirement for accurate and reliable data, the development of appropriate algorithms, and the integration of predictive models into existing workflows. In addition, the finding highlights the growing application of deep learning algorithms for various tasks as one of the major trends. From the analysis of the state-of-the-art in predictive techniques, it is necessary first to survey the landscape of existing predictive analytics approaches and their methods to employ in oil and gas prediction.

1. Introduction

The oil and gas industry affects the global economy and significantly impacts national development. Therefore, the question of how to predict its production rate for subsequent predicting a country's plans by more realistic rules and values [1] by statistical analysis, machine learning, and data mining has gained increase in recent years. Prediction in the oil and gas industry in recent years with the use of oil and gas industry uses machine learning and data analytics to optimize offshore

* *Corresponding author.*
E-mail address: marina998@uitm.edu.my

https://doi.org/10.37934/araset.50.2.260278

and onshore processes. It includes pipeline exploration, drilling, and time-series production. The challenges initiated with traditional methods for forecasting the operational parameters are identified, and case studies are associated with performance optimization using predictive models [2]. Predictive methods are used to analyse and predict various factors, including reservoir performance, drilling efficiency, production rates, equipment failure, and environmental risks. For instance, predicting the long-term production performance and estimated ultimate recovery in unconventional wells has always been challenging [3,4]. In addition, monitoring is limited, and excursions may miss abnormal annulus pressure behaviour within the design envelope in wells.

To overcome this event, a modelling workflow that combines novel deep learning techniques with statistical analysis to create online models that predict asset failures and alert on abnormal behaviour like abrupt pressure build-up in producer and water injection wells' A-Annulus. The model uses autoencoder architecture to learn wells' behaviour during regular operation and alerts when it detects abnormal behaviour [5]. Big data and analytics play a significant role in expediting asset management workflows within the oil and gas industry. Gas compression systems drive process flow and maintain asset uptime in many assets. Due to their complex assembly and high-speed moving parts, compression systems fail frequently. Thus, monitoring and predicting a compression system's operational status improves safety, reliability, and system downtime [6]. A reliable and accurate prediction model in the oil and gas industry would assist in business decision-making. Limited data and machine learning can improve maintenance, production, and asset efficiency [7,8]. Using predictive methods in oil and gas operations has several advantages, such as reducing costs, maximizing production, and minimizing environmental impact.

Furthermore, predictive methods can help companies make informed decisions based on data-driven insights. This significant review explores the recent advances in predictive methods in the oil and gas industry. Specifically, this review will focus on the comparative analysis of different predictive models and their effectiveness in the oil and gas industry. Furthermore, this review will highlight the challenges and limitations of using predictive methods. The research progress and the prospect of machine learning methods on corrosion prediction of oil and gas pipelines. Oil and gas pipeline systems transport oil and natural gas, and accurate corrosion prediction affects pipe material selection [25] and remaining functional life prediction [14]. Machine learning can improve corrosion prediction and control by overcoming mathematical model limitations. Another perspective is that avoiding gas and oil production requires hydrate formation. Several empirical models can predict hydrate formation but depend on system geometries and fluid characteristics [17].

2. Material and Methods

A systematic literature review (SLR) method was used to gather pertinent data and insights from various sources for this paper on predictive models in oil and gas. The SLR method entails thoroughly searching databases, journals, and other pertinent sources to find all relevant studies and publications on a specific topic. Existing research on applying predictive models in the oil and gas industry was found and analysed using the SLR method. To give a thorough overview of the state of predictive modelling in today's market, the research was examined for its applicability, quality, and contribution to the field. The SLR method is a strict and methodical approach that ensures all relevant research is found and examined. It is beneficial in fields such as oil and gas, where a large and constantly evolving body of research exists. Using the SLR method, this paper provided a thorough and reliable analysis of the current state of predictive modelling in the oil and gas industry.

This paper provides a comprehensive review based on advanced searching related to predictive analytics in oil and gas. Advanced evaluation is one of the most critical discussions at the moment.

Thus, a systematic flow method was used in this work. A protocol or plan with clearly stated criteria before a review is referred to as a structured review [9], which is a method for strategically identifying patterns, trends, and critical evaluations [10]. The review technique comprised four steps for choosing numerous relevant papers for this study. This study used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology [11]. It is a framework created to illustrate the flow of information during the various stages of a systematic review, as shown in Figure 1. The first step in writing this comprehensive literature review was identifying research items that may be relevant to the research question. The total number of searched papers was then screened. The third step is the eligibility of each paper based on its abstract was evaluated. Finally, the scientific literature was reviewed and summarized to identify, select, and evaluate breast cancer classification techniques. Subsequently, additional research directions to address the raised concerns were recommended. In this study, the best practice method was used to conduct the comprehensive literature review, and the publication rules provided essential information to help the researchers evaluate the accuracy of the review. Furthermore, an investigation for the systematic analysis of the various studies is considered within this review. The Web of Science (WOS), Scopus, and ScienceDirect databases examined the studied methodologies.

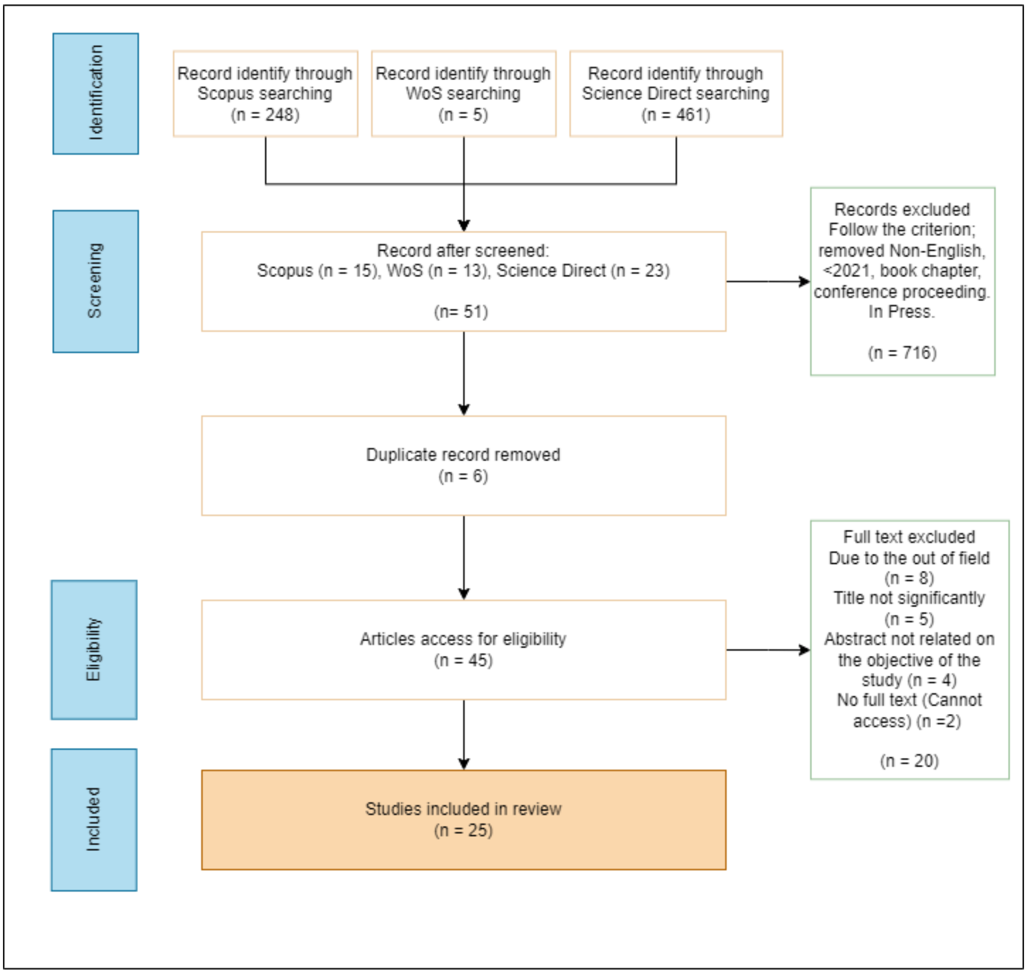

Fig. 1. Flow diagram of the proposed search study

2.1 Identification

In choosing several appropriate papers for this report, the systematic review consists of four main phases. The first step is keyword identification and the quest for linked, similar terms based on the thesaurus, dictionaries, encyclopaedia, and previous studies. Accordingly, after all the relevant keywords were decided, search strings on Scopus, WoS, and ScienceDirect (see Table 1) database have been created. Selected literature articles were identified to determine literature studies on using predictive methods in the oil and gas field. Keywords such as "*machine learning*" OR "*predictive method**" OR "*deep learning*" OR "predictive model*") AND "*oil and gas*" AND "**SHORE*" " were used. The year factors were restricted to 2021 and 2022 to reach all the related recent studies. The Scopus, WoS, and ScienceDirect databases were utilized during the execution of the literature search that was carried out. The preliminary investigation revealed 248 articles from Scopus, 58 articles from WoS, and 461 articles overall, as shown in Table 1. The current research successfully retrieved 767 papers from both databases during the first step of conducting a systematic review.

Table 1
Comparative analysis of predictive model in oil and gas sub-field

Database	Search Strings	Findings
Scopus	TITLE-ABS-KEY (("machine learning" OR "predictive method*" OR "deep learning" OR "predictive model*") AND "oil and gas" AND "*SHORE") AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021)) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j")) Date of access: May 2023	Result = 248 Articles
WoS	*("machine learning"* OR *"predictive method*"* OR *"deep learning"* OR *"predictive model*")* AND *"oil and gas"* AND *"*SHORE")* (Topic) and 2022 or 2021 (Publication Years) and Article (Document Types) and English (Languages) Date of access: May 2023	Result = 58 Articles
Science Direct	*("machine learning"* OR *"predictive method*"* OR *"deep learning" OR "predictive model*")* *AND "oil and gas"* AND *"*SHORE")* Date of access: May 2023	Result = 461 Articles

2.2 Screening

Screening examines relevant research items for content that matches the predefined research question(s). Predictive methods and deep learning in oil and gas are used to select research items in the screening phase. This step removes duplicate papers from the searched list. The first screening eliminated 716 publications, while the second examined six papers based on this study's exclusion and inclusion criteria (see Table 2). Research papers were initially selected as the main criteria due to their ability to offer practical advice. The aforementioned sources encompass a range of academic materials such as reviews, meta-syntheses, meta-analyses, books, book series, chapters, and conference proceedings that were not incorporated in the most recent study. English publications were reviewed and covered in the years 2021 and 2022. Duplication rejected four publications. However, 25 articles were removed because of their premature results and did not discuss predictive analytics in oil and gas. Some of the articles were also incomplete, or the complete articles were not readily accessible, as they contained broken links and exhibited overlapping content.

2.3 Eligibility

The eligibility level corresponds to the third level and contains 45 articles. At this point, the titles of all the articles and the key text were given a thorough examination to check that the inclusion criteria had been satisfied and that the articles were suitable for the research objectives of the current study. All article titles and key text were carefully examined at this point. The selection criterion for searching is in Table 2. As a result, 20 papers were eliminated because, according to empirical data, their titles and abstracts did not significantly relate to the study's goal.

Table 2
The selection criteria for searching

Criterion	Inclusion	Exclusion
Language	English	Non-English
Timeline	2021 – 2022	< 2021
Literature Type	Journal (Article)	Conference, Book, Review
Publication Stage	Final	In Press
Subject Area	Computer Science & Engineering	Besides Computer Science and Engineering/ Others

A further inclusion criterion was that the studies had to be in computer science and engineering. This helped narrow this review to predictive methods and models in oil and gas. Furthermore, sub-field and predictive analytics methods aided in the extraction, and the exclusion criteria excluded articles that focused on different contexts, such as general oil and gas terms and studies not specifically about predictive analytics in offshore and onshore oil and gas.

Figure 2 and Figure 3 show the distribution histograms for the inputs and outputs included in the corrosion rate prediction, respectively [25]. The four input variables are Temperature, C, Wall shear stress (gas phase), Pa, Pressure, kPa, and pH.

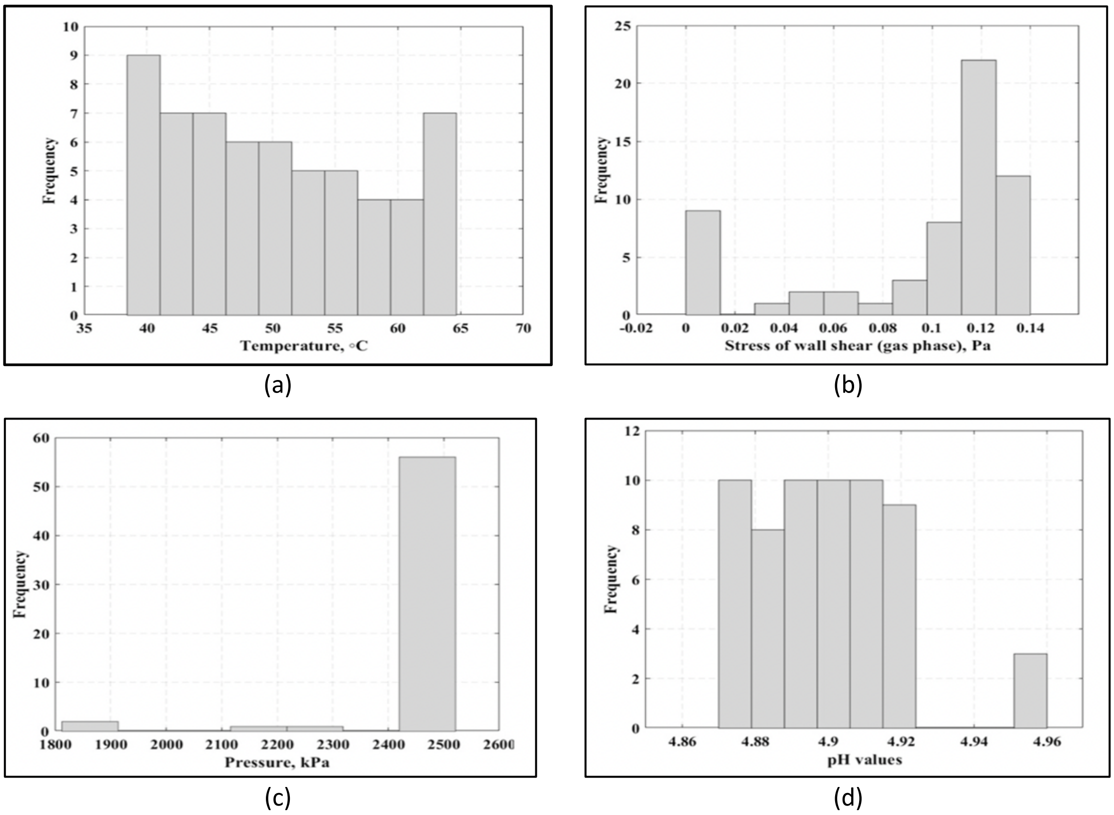

Fig. 2. Example of distribution histograms for input variables (a) Temperature, ◦C (b) Wall shear stress (gas phase), Pa (c) Pressure, kPa (d) pH values

Figure 3 illustrates an output data distribution of corrosion rate. To authenticate the concerns and guarantee the clarity significance, and suitability of the subthemes, it is imperative to conduct an expert review. Mohamad Taufik Mohd Sallehud-din from PETRONAS Research Sdn Bhd, Malaysia was chosen as an oil and gas processes expert. Lastly,25 articles have been made accessible for review.

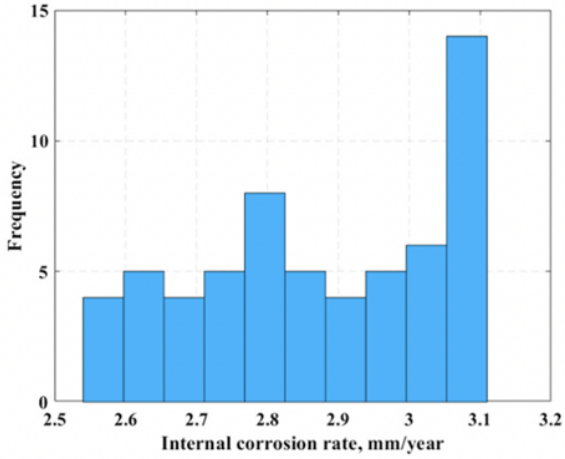

Fig. 3. Example of distribution histograms of the output variable

2.4 Data Extraction and Analysis

In this study, an integrative analysis was employed as one of the assessment strategies to examine and synthesize numerous research designs (quantitative, qualitative, and mixed methods). The goal of the expert study was to identify relevant topics and subtopics. The data-collecting stage was the initial step in the theme's development. As depicted in Figure 1, the authors meticulously analysed 26 publications for assertions or material relevant to the present study's topics. The authors next assess predictive analytics in oil and gas throughout the identifying and establishing significant groupings in the second stage. The two key topics that evolved from the method are the predictive method and the model used. From this point forward, the authors continued each established subject, along with any themes, notions, or ideas. The writer cooperated with other co-authors to create themes depending on the evidence in the context of this research. A log was kept throughout the data analysis process to record any analyses, views, riddles, or other thoughts pertinent to the data interpretation. Finally, the authors contrasted the results to see any inconsistencies in the theme design process. It is worth mentioning that if there are any discrepancies among the concepts, the authors discuss them among themselves. Eventually, the produced themes were tweaked to ensure that they were consistent. Experts conducted the analysis, one specializing in oil and gas related to predictive analytics to establish the validity of the problems. By establishing domain validity, the expert review phase helps ensure each sub-theme's clarity, importance, and suitability. Based on comments and professional judgments, the writer makes amendments to his or her judgment.

3. Results and Findings

Data analysis, particularly prediction analytics, has been used in offshore and onshore oil and gas research as the oil and gas industry has gained popularity. As a result, time and money spent on platform operation and crude oil prediction have been reduced. To help with prediction, both conventional machine learning techniques and cutting-edge techniques like Long Short-Term Memory (LSTM) and Convolution Neural Network (CNN) were employed. The models provide a specific result based on the sub-fields and their data. However, to demonstrate that the model can be helpful to researchers working in this field, a comparative analysis of it using the most direct comparison achievable is required. In addition, the research investigates the application of a predictive method or model with oil and gas operations and data in the oil and gas industry. This study evaluates the use of a predictive algorithm associated with various types of data and the proposed predictive methods. Among popular techniques for predictive analytics are Artificial Neural Network (ANN), LSTM, and CNN. Results, advantages, and shortcomings are briefly added. Twenty-five (25) articles were extracted and analysed using the search technique.

3.1 Comparative Analysis of Predictive Models

This theme can compare the effectiveness and accuracy of different predictive models used in the oil and gas industry, such as machine learning and deep learning. Comparative analysis of predictive models involves evaluating and comparing the performance of different models in making accurate predictions. Examples of predictive models are predicting drilling efficiency, wave height, CO2 degradation prediction, remaining useful life (RUL), and estimating crude oil prices. To conduct a comparative analysis of predictive models in oil and gas operations, researchers typically select models to evaluate and compare their performance using relevant performance metrics. These metrics may include accuracy, precision, Root Mean Square Error (RMSE), Mean Square Error (MAE), and Mean Absolute Percentage Error (MAPE).

The comparative analysis may involve using different models, such as machine learning and deep learning models, and evaluating their performance based on various input data sources, such as drilling data, carrion, crack, and time series data. Researchers may also consider different model training and validation techniques, such as training, testing accuracy, and statistical tests, to ensure the robustness and reliability of the results. The comparative analysis aims to identify the most effective and accurate predictive model(s) for a given oil and gas operation and to understand each model's strengths and weaknesses. This information can improve the accuracy and effectiveness of predictive models in oil and gas operations, ultimately leading to improved efficiency, cost savings, and environmental impact reduction. Many researchers have contributed to predictive analytics in oil and gas. A summary of methods and outcomes of comparisons between studies. Table 3 summarizes machine learning and deep learning research on oil and gas operations offshore and onshore.

Table 3
Comparative analysis of predictive model in oil and gas sub-field

Reference	Sub-field	Dataset	Methodology	Results	Advantages/ Shortcomings
[12]	Offshore platform operation	Case Study 1: Natural Gas treatment plant data Case Study 2: Sea water injection pump for oil.	Pre-processing of the sensor data and LSTM and CNN+LSTM model construction and evaluation. Type: CNN+LSTM	Case 1: LSTM precision 83.50% Case 2: CNN+LSTM precision 99.00	The proposed CNN+LSTM for degradation prediction provides better performance. Prediction of elevated levels of CO_2 obtain less performance.

[13]	RUL (U.S. federal waters of the Gulf of Mexico of natural and engineered offshore system)	Gradient boosted regression tree (GBRT) and ANN is constructed to quantify operating platforms' remaining useful life. Type: ANN	GBRT and ANN: Accuracy: 95.00–97.00%.	Prediction of the RUL of operating platforms would assist in localizing maintenance strategies that will lead to preventing operational and environmental risks while maintaining energy production.
[14]	RUL	A multivariate CNN to predict early failure behaviour in sensor-instrumented tribosystem. Type: CNN	Training accuracy: 99.00% Testing accuracy: 95%.	Predictions of RUL on the remaining beneficial lifetime of ball-bearing-type components would help tribological machine elements trigger automated maintenance.
[15]	Two oil wells in an offshore oil field in the South China Sea	Feature engineering, parameter optimization, three data-driven models, Back Propagation (BP) neural network, LSTM, and RF. Type: LSTM	LSTM model shows the best performance. Accuracy (MAE): LSTM 1.95 LSTM 3.9 % BP 12.1% RF moderate	LSTM model has immense potential in applying virtual flow meters that can work for intelligent green oil and gas engineering.
[16]	Iceberg-seabed interaction	Eight extra tree regression (ETR) models. Evaluation by randomly split 70:30. Type: Regression	The performance of ETR models is better than decision tree regression (DTR), random forest regression (RFR), and Gradient boosting regression (GBR) algorithms.	All ETR models perform better in simulating the ice keel seabed interaction in clay.

[17]		Time-series data (200 rpm, 600 rpm)	Deep learning models (Dense, LSTM, Gated Recurrent Unit (GRU), Attention Residual LSTM (ARLSTM), layers, and dropout iterative transfer learning. Type: LSTM	ARLSTM obtains the smallest error compared to Dense, LSTM, and GRU	Using sequential time series data, ARLSTM predicts good kinetic characteristics during hydrate formation in real time. It identifies gas production pipeline risks.
[18]		Shale gas well production.	Exponential smoothing method, Autoregressive Integrated Moving Average (ARIMA), LSTM, Arps, stretched exponential decline (SEPD), and Duong. Type: LSTM	LSTM model predicted short- and long-term events with scientific accuracy.	Deep learning in the petroleum industry and unconventional hydrocarbon production prediction.
[19]		Real-time drilling geodata.	Pre-processing, ANN modelling. Type: ANN	The accuracy of predicting drilling is identified.	ANN can reduce unproductive time spent on eliminating stuck pipes, mud losses, and gas, oil, and water and prevent complications in well drilling for a zero-carbon footprint in the environmentally friendly drilling of wells on the land and offshore.
[20]	Offshore, pipelines	Crack propagating in offshore piping (Input Variable: crack depth and half crack length.	Adaptive Gaussian Process Regression Model (AGPRM) Type: Regression	AGPRM Squared correlation coefficient 0.535 RMSE 0.545 GPR 0.993 MLP 0.992 Water depth at low IP increases the pipeline's maximum Von-Mises Stress (MVMS).	A significant finding for GPR and MLP on the prediction of stress intensity factor (SIF) to assess the remaining fatigue life (RFL) of offshore pipelines.

[21]		Prediction Models: Genetic Algorithm (GA) based Back Propagation neural network (GA-BP), Radial Basis Function (RBF), and Support Vector Machine (SVM) Models used for forecasting pipeline equilibrium scour depth. Type: ANN+Opt	GA-BP model better RBF and SVM. GA-BP: Correlation coefficient lowest. RMSE lowest.	Sensitivity analysis shows that Froude number (Fr) best predicts pipeline scouring
[22]	300 samples of maximum pitting corrosion depth.	Integrated Singular Spectrum Analysis (SSA) and LSTM. Type: Decompose+LSTM	SSA-LSTM model: MAE 8.84% RMSE 0.06 MSE 0.36% MAPE 9.58%	Balanced SSA-LSTM integration. LSTM assesses pipeline corrosion. Global search and fast convergence help SSA optimize hyperparameters. The model can digitalize subsea process system safety online.
[23]	Niger Delta Area of Nigeria field Egua-1 (NDANE-1flow line).	Mechanistic Modelling approach with DeWaard Milliams' and Applied. Type: Other Method	Applied-Model RMSE 0.022 MAE 0.018 SI 0.258 R2 0.915 Measurement Dewaard Millaims Model: RMSE 0.052 MAE 0.0420 SI 0.495 R2 0.533	Global oil and gas companies could predict corrosion rates using an applied model on multiphase flow parameters affecting pipeline corrosion.
[24]	Corroded pipeline data.	Latin Hypercube Sampling method along with Simulated Annealing. Type: Other Method	Corroded pipelines may have less life and capacity because the deterministic analysis conservatively predicts the internal pressure corresponding to the failure limit state.	Lack of data necessitates investigating pipeline mechanical and geometric properties and corrosion defects on failure strength and remaining life.

[25]		Corrosion	Four ensemble approaches: RF, adaptive boosting, Gradient boosting regression tree, and extreme gradient boosting. Evaluation with k-fold cross-validation. Type: Ensemble	The extreme gradient boosting model outperforms other approaches. RMSE for internal corrosion rate 0.031 mm/y Performance index, PI = 0.61.	All ensemble models perform well, but extreme gradient boosting is the most practical for reducing internal corrosion rates in oil and gas pipelines. A sensitivity analysis using feature importance criteria showed the strongest corrosion rate dependency on temperature and pressure, along with CO2.
[26]	Offshore, sub-sea	Satellite altimetry	Application of ANN, SVM, and Random Forest. Type: Ensemble	RF performs better than ANN and SVM for wave period prediction, ANN performs better than RF and SVM wave height prediction.	Prediction of wave height and period are expected to test with other ML methods.
[27]		Ice-scoured.	Fifteen Self-adaptive Extreme Learning Machine (SAELM) models with 70: 30 data splits. Type: ELM	SAELM models are the most influencing input parameters in the estimation of the sub-gouge clay characteristics.	SAELM models are good at sub-gouge clay estimation.

[28]	Onshore, hydrocarbon	Hydrocarbon accumulation assemblages.	Regression algorithm multiple ANN, multiple linear regression, random forest and support vector machine methods. Type: ANN	ANN outperforms multiple linear regression, RF, and SVM. ANN Average relative errors: No. 1 23.33%, No. 2: 24.93% No. 3: 26.38%, No. 4: 22.60%,	ANN method was significantly in forecasting hydrocarbon accumulation better than the traditional model.
[29]	Onshore, crude oil	West Texas Intermediate (WTI) and Brent Crude Oil Time Series COTS.	LSTM+Henry gas solubility optimization (CHGSO) technique. Evaluation by Exponential moving average (EMA), Simple Moving average (SMA) and Kaufman's adaptive moving average (KAMA), commodity channel index (CCI), rate of change (ROC), and relative strength index (RSI), and the volatility indicators such as average true range (ATR), volatility ratio (VR) and highest high-lowest low (HHLL). Type: LSTM+Opt	Extraction on the CHGSO algorithm with the logistic chaotic map is better than EMA, SMA, Kaufman's adapt, and Kaufman's adaptive moving average (KAMA). CCI, ROC, RSI, ATR, VR and HHLL.	LSTM+CHGSO has the ability to estimate crude oil prices with high accuracy and overcome issues of chaotic and nonlinear characteristics on trend, momentum, and volatility technical indicators. Theil's U and MAPE are objective functions that are workable.
[30]		Fourier-transform infrared spectroscopy (FTIR).	Two approaches with data reduction strategy: Principal component analysis (PCA) and the support vector regression (SVR), (PCA-SVR) and Autoencoder and the SVR (Auto-SVR). Type: Regression	Both approaches achieve satisfactory prediction accuracy.	The strategy of PCA by extracting the high-dimension FTIR data from lower-dimensional and autoencoder able to learn new representations for the dimensionality reduction of the FTIR data make the SVR good in crude oil properties prediction.

[31]	West Texas Intermediate (WTI) oil price (January 02, 2009, to November 25, 2019)	Time series prediction of crude oil futures price using Lyapunov exponent, Bootstrap technique to verify the robustness of the largest Lyapunov, Construct eight models with ANN technology and Chaos theory. Type:Ensemble+ANN	EMD-LR-CHAOS model appeared to be the best prediction model among the eight models.	The chaotic intrinsic formation mechanism (chaotic) model is acceptable and can be improved because it would give investors a high return with little risk.
[32]	WTI (1991 to 2021)	Logistic Regression, Decision Tree, Random Forest, AdaBoost, and XgBoost, use DeLong statistical test procedures and Shapley Additive explanations values to support model evaluation and interpretability. Type: Ensemble	All models can predict the trend of WTI crude oil prices. DeLong statistical test procedures to accurately compare machine learning models' performance.	The findings are significant for policymakers, businesses, investors, and long-term energy-based economic development.
[33]	Traditional economic data and Google search volume index (GSVI)	K-means＋Kernel Principal Component Analysis (KPCA)＋Kernel extreme learning machine (KELM) based on a "divide and conquer" strategy. Type: Integrated ELM	The "divide and conquer" strategy improves forecasting performance. Data level forecasting accuracy: GSVI data performs better than economic data in level.	K-Means＋KPCA＋KELM on both datasets can improve the accuracy of monthly crude oil price forecasting.

[34]		Brent, WTI (January 4, 2010 - July31, 2 020)	Hybrid Wavelet Transform (WT), Bidirectional long short-term memory network (BiLSTM)-Attention-CNN (BLAC) (WT-BLAC). Type: Decompose+CNN+LSTM	Results WT-BLAC: WTI R^2 0.9663 RMSE 2.2518 MAPE 1.1780 MAE 2.6261 Brent $R^2$0.9900 RMSE 1.2936 MAPE 1.8027 MAE 0.8894	WT-BLAC outperforms other models and has potential for government agencies, investors, and related businesses.
[35]		Oil formation (Egyptian)	ANN model. Type: ANN	ANN model R2 0.974 ARE 0.0017%, AARE 2.13%	The ANN model outperforms local Egyptian empirical correlations and all other global models in sustainability.
[36]	Onshore, Operation platform	Time series of the motion.	DL uses 6-DOF motions to generate six standalone subnets in about 15 seconds (one wave cycle). The motions were reconstructed over 3 hours. Dropout probability, number of RNN layers, number of fully connected layers, and number of neurons in each layer are all hyperparameters. Type: DL	DL works well for offshore engineering.	Predicting the motions of a floating offshore structure could help a motion compensation system and alert motion-sensitive operations.

Predictive analytics is a broad field with many sub-fields or domains of focus in the oil and gas industry. This paper focuses on the main fields of offshore and onshore predictive analytics solution strategy. A few sub-fields in this field can be classified based on the specific problems of the industry. Figure 4 depicts the total number of sub-field research on predictive analytics in the oil and gas industry, focusing on offshore and onshore fields. Platform operation, pipelines, crude oil, and offshore and onshore drilling are instances of such sub-fields. The number of articles published on the most recent predictive analytics method is nearly identical for platform operations, pipelines, and crude oil. The most popular, however, is platform operation.

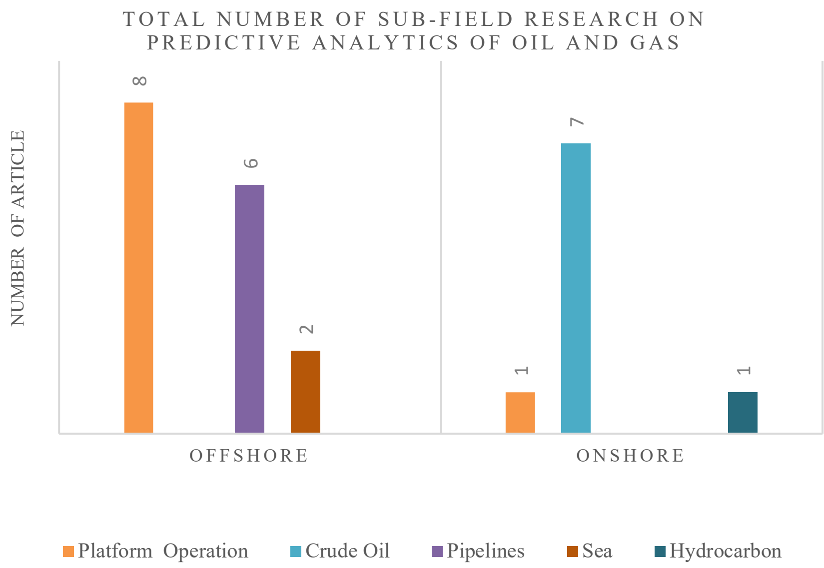

Fig. 4. Total number of sub-field research on predictive analytics of oil and gas based on offshore and onshore field

The best predictive analytics methods vary according to the problem domain, data, and field state. The predictive analytics method has evolved from traditional ML to modern ML. Modern DL typically refers to methods such as CNN and LSTM. Figure 5 depicts the number of best predictive analytics methods. Regarding the oil and gas sub-field, LSTM and its variants have demonstrated the most usable methods. Aside from that, the ensemble strategy is becoming more attractive. All ensemble models, for example, RF, adaptive boosting, Gradient boosting, a regression tree, and extreme gradient boosting, perform well in decreasing corrosion rates in oil and gas pipelines [25] and satellite altimetry, as seen in Table 1. However, CNN, regression, and extreme learning machines (ELM) have shown less priority.

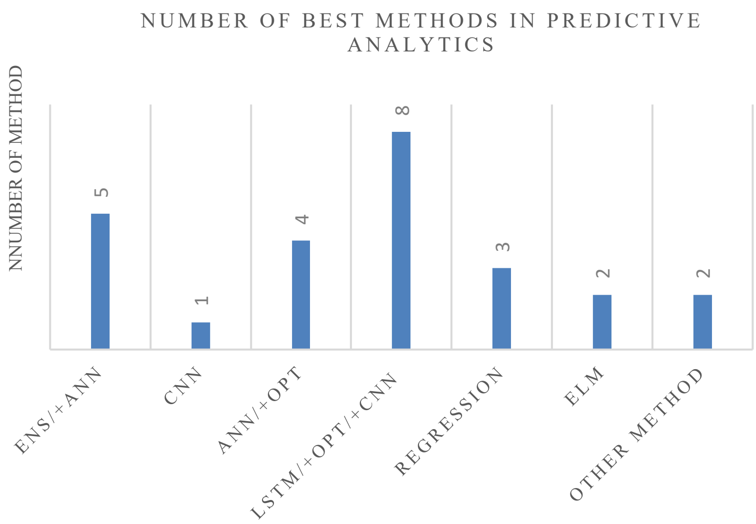

Fig. 5. Number of best methods in predictive analytics

One of the possible explanations for these findings is that there have been several prior attempts at predictive analytics in the oil and gas industry. LSTM is currently one of the best approaches for predictive analytics on most sub-fields. Previous researchers have studied several techniques related to LSTM, such as LSTM with optimization (LSTM+Opt) and CNN with LSTM (CNN+LSTM). CNN+LSTM, for example, is used to solve two case studies of degradation prediction in offshore operation platforms for natural gas treatment plants and seawater injection pumps for oil [12]. The precision achieved in the study is 99.00%, according to the results. Compared to a single LSTM, it performs better, with a 15.5% improvement in precision. LSTM performance has also been reported in good time series [17,18]. Yang *et al.,* [18] used the LSTM model to predict short- and long-term shale gas well production events. It outperformed the ARIMA Arps, Stretched exponential Decline (SEPD), and Duong methods.

Furthermore, Song *et al.,* [15] reported that the LSTM with feature engineering and parameter optimization (LSTM+FE+Opt) model has the lowest MAE value when compared to BPNN, LSTM, and Random Forest (RF) [13] calculated accuracy of 97% in predicting the RUL in an operating platform using Gradient boosted regression tree (GBRT) and Artificial Neural Network (ANN). Almost identical results were obtained using CNN on training and testing accuracy of approximately 99.00% and 95.00%, respectively. The models' performance would reduce operational and environmental risk while maintaining energy production by forecasting RUL operation.

Another research to consider is the estimation of crude oil prices. Karasu and Altan developed the LSTM+Henry gas solubility optimization (CHGSO) technique to estimate crude oil prices accurately [34]. They used West Texas Intermediate (WTI) and Brent Crude Oil Time Series COTS datasets. Hybrid Wavelet Transform (WT), Bidirectional Long Short-Term Memory Network (BiLSTM)-Attention-CNN (BLAC) (WT-BLAC). WT-BLAC performs well for WTI data with R2, RMSE, MAPE, and MAE of 0.9663, 2.2518, 1.1780, and 2.6261, respectively. Furthermore, Guliyev & Mustafayev [32] used a similar dataset evaluated using ensemble and hybrid Ensemble+ANN but with a different range of time series data. The models have acceptable and significant interpretability in time series prediction of crude oil futures prices.

4. Discussion

The review on predictive methods in oil and gas provides a comprehensive overview of the latest research in the field, including various applications of machine learning and deep learning techniques. One of the key trends observed in the review is the increasing use of deep learning algorithms for a wide range of tasks. It provides an investigation into the current state of the art in terms of predictive methods for oil and gas operations, including a comparison of the efficacy of various predictive methods. These methods can be used to develop models that predict factors such as reservoir performance, drilling efficiency, and production rates. For example, predictive models can be developed to analyse data from seismic surveys to predict the location and characteristics of oil and gas reservoirs. Similarly, machine learning algorithms can be used to analyse data from drilling and production operations to identify patterns and trends that can be used to optimize operations and reduce costs.

Other predictive methods for onshore oil and gas operations include using data analytics to predict equipment failure and maintenance needs, as well as predicting environmental risks and impacts. Predictive methods utilization in onshore oil and gas operations can help companies improve efficiency, reduce costs, and maximize production while minimizing environmental impact. The review also identifies several challenges that must be addressed to fully realize the potential of predictive methods in the oil and gas industry. These challenges include the need for large amounts of high-quality data, more interpretable models, and effective collaboration between data scientists and domain experts. Overall, the review highlights the potential of predictive methods to improve operational efficiency, reduce costs, and enhance safety in the oil and gas industry. However, more research is needed to address the challenges and ensure these methods can be successfully applied in real-world scenarios.

The application of predictive techniques in the oil and gas industry involves implementing diverse deep learning methods and their variations within the exploration and production procedures. Implementing these techniques has dramatically improved the operational effectiveness of the oil and gas sector. However, it is crucial to carefully evaluate certain factors, specifically regarding the precision and reliability of predictive models. Challenges manifest as a result of data incompleteness, inconsistency, or noise originating from engineering procedures, such as estimating crude oil prices and predicting pipeline failures. To ensure the viability and acceptability of the data acquired from real-time or near real-time sensors, it is imperative to perform data pre-processing. Therefore, integrating automated pre-processing techniques using suitable methodologies should be included in developing advanced predictive methods. Another facet to consider is the preparedness of the data [37] and further analysis for identifying the optimum value of the parameters for oil and gas business to stay competitive [38]. Numerous oil and gas companies possess extensive legacy systems and data silos, presenting challenges in integrating data from diverse sources into a cohesive predictive model.

5. Conclusion

The review compares the effectiveness of various predictive methods for offshore and onshore oil and gas operations, such as machine learning and deep learning. The article investigates the benefits and drawbacks of each method and provides insight into which method is most effective for different oil and gas operations. It can also recommend best practices in using predictive methods in the industry. It also discusses how predictive models have been developed and used to improve platform operation and time-series crude oil, pipeline corrosion, and drilling efficiency to increase production rate and lower costs. It may also provide information on best practices for using predictive methods in the industry, such as integrating predictive models with real-time data analytics. Because of the complexity of the production process and the scarcity of data, predictive modelling in the oil and gas industry faces several challenges. However, by carefully considering these challenges and being willing to invest in the necessary resources, the oil and gas industry can overcome them and leverage the power of predictive models to optimize production, reduce costs, and remain competitive in an increasingly competitive market.

Acknowledgment
The authors would like to acknowledge PETRONAS Academia Collaboration Grant, Institute for Big Data Analytics and Artificial Intelligence (IBDAAI) and Universiti Teknologi MARA for the financial support provided to this research project.

References
[1] Majed, Hadeer, Samaher Al-Janabi, and Saif Mahmood. "Data Science for Genomics (GSK-XGBoost) for Prediction Six Types of Gas Based on Intelligent Analytics." In *2022 22nd International Conference on Computational Science and Its Applications (ICCSA)*, pp. 28-34. IEEE, 2022. https://doi.org/10.1109/ICCSA57511.2022.00015

[2] Pandey, Rakesh Kumar, Anil Kumar Dahiya, and Ajay Mandal. "Identifying applications of machine learning and data analytics based approaches for optimization of upstream petroleum operations." *Energy Technology* 9, no. 1 (2021): 2000749. https://doi.org/10.1002/ente.202000749

[3] Alarifi, Sulaiman A., and Jennifer Miskimins. "A new approach to estimating ultimate recovery for multistage hydraulically fractured horizontal wells by utilizing completion parameters using machine learning." *SPE Production & Operations* 36, no. 03 (2021): 468-483. https://doi.org/10.2118/204470-PA

[4] Al Dhaif, Redha, Ahmed Farid Ibrahim, and Salaheldin Elkatatny. "Prediction of surface oil rates for volatile oil and gas condensate reservoirs using artificial intelligence techniques." *Journal of energy resources technology* 144, no. 3 (2022): 033001. https://doi.org/10.1115/1.4051298

[5] Jain, A., A. Morgenthal, M. Aman, M. Horton, and S. Khan. "Creating an Auto-Encoder Based Predictive Maintenance Tool for Offshore Annulus Wells." In *SPE Annual Technical Conference and Exhibition?*, p. D022S083R001. SPE, 2022. https://doi.org/10.2118/210220-MS

[6] Cao, Kai, Zhenduo Zhang, Ying Li, Wenbo Zheng, and Ming Xie. "Ship fuel sulfur content prediction based on convolutional neural network and ultraviolet camera images." *Environmental Pollution* 273 (2021): 116501. https://doi.org/10.1016/j.envpol.2021.116501

[7] Cadei, Luca, Andrea Corneo, Diletta Milana, Danilo Loffreno, Lorenzo Lancia, Marco Montini, Gianmarco Rossi, Elisabetta Purlalli, Piero Fier, and Francesco Carducci. "Advanced Analytics for Predictive Maintenance with Limited Data: Exploring the Fouling Problem in Heat Exchanging Equipment." In *Abu Dhabi International Petroleum Exhibition and Conference*, p. D032S204R002. SPE, 2019. https://doi.org/10.2118/197355-MS

[8] Cadei, Luca, Gianmarco Rossi, Lorenzo Lancia, Danilo Loffreno, Andrea Corneo, Diletta Milana, Marco Montini *et al.,* "Digital Lighthouse: A Scalable Model for Digital Transformation in Oil & Gas." In *SPE EOR Conference at Oil and Gas West Asia*, p. D031S023R001. SPE, 2022. https://doi.org/10.2118/200149-MS

[9] Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer *et al.,* "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." *bmj* 372 (2021).

[10] Rethlefsen, Melissa L., Shona Kirtley, Siw Waffenschmidt, Ana Patricia Ayala, David Moher, Matthew J. Page, and Jonathan B. Koffel. "PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews." *Systematic reviews* 10 (2021): 1-19. https://doi.org/10.1186/s13643-020-01542-z

[11] Moher, David, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and T. PRISMA Group*. "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." *Annals of internal medicine* 151, no. 4 (2009): 264-269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

[12] Figueroa Barraza, Joaquin, Luis Guarda Bräuning, Ruben Benites Perez, Carlos Bittencourt Morais, Marcelo Ramos Martins, and Enrique Lopez Droguett. "Deep learning health state prognostics of physical assets in the Oil and Gas industry." *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 236, no. 4 (2022): 598-616. https://doi.org/10.1177/1748006X20976817

[13] Dyer, Alec S., Dakota Zaengle, Jake R. Nelson, Rodrigo Duran, Madison Wenzlick, Patrick C. Wingo, Jennifer R. Bauer, Kelly Rose, and Lucy Romeo. "Applied machine learning model comparison: Predicting offshore platform integrity with gradient boosting algorithms and neural networks." *Marine Structures* 83 (2022): 103152. https://doi.org/10.1016/j.marstruc.2021.103152

[14] Desai, Prathamesh S., Victoria Granja, and C. Fred Higgs III. "Lifetime prediction using a tribology-aware, deep learning-based digital twin of ball bearing-like tribosystems in oil and gas." *Processes* 9, no. 6 (2021): 922. https://doi.org/10.3390/pr9060922

[15] Song, Jingqi, Yuanjie Zheng, Jing Wang, Muhammad Zakir Ullah, Xuecheng Li, Zhenxing Zou, and Guocheng Ding. "Multi-feature deep information bottleneck network for breast cancer classification in contrast enhanced spectral mammography." *Pattern Recognition* 131 (2022): 108858. https://doi.org/10.1016/j.patcog.2022.108858

[16] Azimi, Hamed, and Hodjat Shiri. "Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine." *Ocean Engineering* 239 (2021): 109795. https://doi.org/10.1016/j.oceaneng.2021.109795

[17] Lee, Nayoung, Hyunho Kim, JongYeon Jung, Ki-Heum Park, Praveen Linga, and Yutaek Seo. "Time series prediction of hydrate dynamics on flow assurance using PCA and Recurrent neural networks with iterative transfer learning." *Chemical Engineering Science* 263 (2022): 118111. https://doi.org/10.1016/j.ces.2022.118111

[18] Yang, Run, Xiangui Liu, Rongze Yu, Zhiming Hu, and Xianggang Duan. "Long short-term memory suggests a model for predicting shale gas production." *Applied Energy* 322 (2022): 119415. https://doi.org/10.1016/j.apenergy.2022.119415

[19] Dmitrievsky, A N, N A Eremin, A D Chernikov, and S O Borozdin. "Intelligent Complication Prevention Systems for Safe Well Construction." *Bezopasnost' Truda v Promyshlennosti* 6 (2022): 7–13. https://doi.org/10.24000/0409-2961-2022-6-7-13

[20] Abyani, Mohsen, Mohammad Reza Bahaari, Mohamad Zarrin, and Mohsen Nasseri. "Predicting failure pressure of the corroded offshore pipelines using an efficient finite element based algorithm and machine learning techniques." *Ocean Engineering* 254 (2022): 111382. https://doi.org/10.1016/j.oceaneng.2022.111382

[21] Hu, Ke, Xinglan Bai, Zhaode Zhang, and Murilo A. Vaz. "Prediction of submarine pipeline equilibrium scour depth based on machine learning applications considering the flow incident angle." *Applied Ocean Research* 112 (2021): 102717. https://doi.org/10.1016/j.apor.2021.102717

[22] Li, Xinhong, Mengmeng Guo, Renren Zhang, and Guoming Chen. "A data-driven prediction model for maximum pitting corrosion depth of subsea oil pipelines using SSA-LSTM approach." *Ocean Engineering* 261 (2022): 112062. https://doi.org/10.1016/j.oceaneng.2022.112062

[23] Obaseki, M., B. N. Nwankwojike, and F. I. Abam. "Diagnostic and prognostic development of a mechanistic model for multiphase flow in oil-gas pipelines." *Journal of King Saud University-Engineering Sciences* 34, no. 8 (2022): 562-570. https://doi.org/10.1016/j.jksues.2020.12.010

[24] Abyani, Mohsen, and Mohammad Reza Bahaari. "A new approach for finite element based reliability evaluation of offshore corroded pipelines." *International Journal of Pressure Vessels and Piping* 193 (2021): 104449. https://doi.org/10.1016/j.ijpvp.2021.104449

[25] Seghier, Mohamed El Amine Ben, Daniel Höche, and Mikhail Zheludkevich. "Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques." *Journal of Natural Gas Science and Engineering* 99 (2022): 104425. https://doi.org/10.1016/j.jngse.2022.104425

[26] Azad, Mohammad, and Md Alhaz Uddin. "Prediction of Offshore Wave at East Coast of Malaysia—A Comparative Study." *Electronics* 11, no. 16 (2022): 2527. https://doi.org/10.3390/electronics11162527

[27] Azimi, Hamed, Hodjat Shiri, and Eduardo Ribeiro Malta. "A non-tuned machine learning method to simulate ice-seabed interaction process in clay." *Journal of Pipeline Science and Engineering* 1, no. 4 (2021): 379-394. https://doi.org/10.1016/j.jpse.2021.08.005

[28] Liu, Guowen, Wangshui Hu, Xiyuan Li, and Binchi Zhang. "The division of oil and gas accumulation assemblage in Sichuan Basin and the construction of favorable accumulation assemblage prediction model." *Energy Reports* 8 (2022): 14716-14725. https://doi.org/10.1016/j.egyr.2022.10.373

[29] Karasu, Seçkin, and Aytaç Altan. "Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization." *Energy* 242 (2022): 122964. https://doi.org/10.1016/j.energy.2021.122964

[30] Yang, Shu-Bo, Jesús Moreira, and Zukui Li. "Predicting crude oil properties using fourier-transform infrared spectroscopy (FTIR) and data-driven methods." *Digital Chemical Engineering* 3 (2022): 100031. https://doi.org/10.1016/j.dche.2022.100031

[31] Yin, Tao, and Yiming Wang. "Predicting the price of WTI crude oil futures using artificial intelligence model with chaos." *Fuel* 316 (2022): 122523. https://doi.org/10.1016/j.fuel.2021.122523

[32] Guliyev, Hasraddin, and Eldayag Mustafayev. "Predicting the changes in the WTI crude oil price dynamics using machine learning models." *Resources Policy* 77 (2022): 102664. https://doi.org/10.1016/j.resourpol.2022.102664

[33] Yang, Yifan, Ju'E. Guo, Shaolong Sun, and Yixin Li. "Forecasting crude oil price with a new hybrid approach and multi-source data." *Engineering Applications of Artificial Intelligence* 101 (2021): 104217. https://doi.org/10.1016/j.engappai.2021.104217

[34] Lin, Yu, Kechi Chen, Xi Zhang, Bin Tan, and Qin Lu. "Forecasting crude oil futures prices using BiLSTM-Attention-CNN model with Wavelet transform." *Applied Soft Computing* 130 (2022): 109723. https://doi.org/10.1016/j.asoc.2022.109723

[35] Gouda, Abdelrahman, and Attia Mahmoud Attia. "Development of a new approach using an artificial neural network for estimating oil formation volume factor at bubble point pressure of Egyptian crude oil." *Journal of King Saud University-Engineering Sciences* (2022).

[36] Guo, Xiaoxian, Xiantao Zhang, Wenyue Lu, Xinliang Tian, and Xin Li. "Real-time prediction of 6-DOF motions of a turret-moored FPSO in harsh sea state." *Ocean Engineering* 265 (2022): 112500. https://doi.org/10.1016/j.oceaneng.2022.112500

[37] Martin, Awaludin, and Hamdani Wahab. "Energy and Thermo-Economic Analysis of Crude Oil Gathering Station and Hydrocarbon Transport." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 97, no. 2 (2022): 146-156. https://doi.org/10.37934/arfmts.97.2.146156

[38] Muniandy, Sumitra, Syuhaida Ismail, and Md Ezamudin Said. "Revenue/cost production sharing contract (psc) fiscal regime on marginal gas fields in Malaysia: Case study." *Progress in Energy and Environment* (2023): 11-18. https://doi.org/10.37934/progee.26.1.1118