Global Journal of Environmental Science and Technology: ISSN-2360-7955
Abbreviated Key Title: Glob. J. Environ. Sci. Technol. ISSN: 2384-5058. Open Access.
Vol. 14(1): Pp. 9-15, January, 2024. DOI: 10.54978

Asset Integrity and Reliability: Data-Driven Non-parametric Test and Machine Learning Approach for Crude Oil Pressure Vesselsa

[1*]Andrew John, [2]Amadi Goodluck Ohahuru, [3]Famous Jacob Ajoboru

[1,2]Department of Welding Engineering and Offshore Technology, Petroleum Training Institute, Effurun, Nigeria.
[3]Department of Mechanical Engineering, Petroleum Training Institute, Effurun, Nigeria.

Corresponding author email: john_a@pti.edu.ng

Abstract

Asset integrity management is a complex task in the oil industry, particularly in surface production facilities. Pressure vessels, crucial for storing and transporting gases and liquids under high pressure, are susceptible to various damage mechanisms. Ensuring their integrity is vital to prevent catastrophic failures. This study aims to develop a multidisciplinary approach to asset integrity management (AIM) by integrating expert knowledge, survey questionnaires, and data analytics (non-parametric and machine learning). Two sets of questionnaires were developed in which the Kendall Coefficient of Concordance (W) ranked the features of AIM and the Random Forest Classifier ranked the NDT techniques. W showed that Corrosion (79), Pressure (84), Temperature limits (87), Vibration (93), Maintenance Strategies (96), Inspection Techniques (106) are the most important 'AIR' parameters while the Phased Array ultrasonic testing, Time of flight Diffraction, Acoustic Emission Testing, Eddy current testing, Pulsed Eddy current are the most important NDT techniques. The study provides a practical framework for controlling and minimizing incidents in oil and gas operations, ultimately contributing to improved safety and efficiency by providing insights into which features are most influential as regards AIR and NDT techniques.

Keywords: Asset Integrity, Data-Driven, Non-parametric Test, Machine Learning

1. INTRODUCTION

Asset integrity and reliability are critical concerns in the oil and gas industry, particularly in the context of crude oil pressure vessels [1]. These vessels are subject to extreme operating conditions, including high pressures and temperatures, which can lead to degradation and failure over time [2]. The consequences of vessel failure can be severe, resulting in costly repairs, production downtime, and potentially catastrophic safety and environmental risks [3]. In recent years, the increasing availability of sensor data and advances in machine learning and data analytics have created new opportunities for improving asset integrity and reliability. By leveraging these technologies, operators can gain deeper insights into the condition and performance of their vessels, enabling more informed maintenance and repair decisions [4][5] [6]. The importance of asset integrity and reliability in the oil and gas industry is well established. Various studies have highlighted the need for effective maintenance and repair strategies to ensure the safe and reliable operation of critical assets like crude oil pressure vessels [7][8]. Traditional approaches to vessel integrity management rely heavily on periodic inspections and manual data collection, which can be time-consuming, costly, and prone to human error [9]. In contrast, data-driven approaches leveraging machine learning and advanced analytics offer significant potential for improving the accuracy and efficiency of vessel integrity assessments [10]. Non-parametric tests, such as the Kolmogorov-Smirnov test and the Mann-Whitney U test, have been widely used in various fields for comparing distributions and identifying patterns in data [11]. In the context of vessel integrity management, these tests can be used to identify anomalies and trends in sensor data that may indicate potential issues with vessel performance or condition [7]. Machine learning algorithms, such as neural networks and decision trees, have also been successfully applied to various problems in the oil and gas industry, including predictive maintenance and asset integrity management [8][10]. These algorithms can be trained on historical data to learn patterns and relationships between different variables, enabling accurate predictions and classifications [12].

Several studies have demonstrated the effectiveness of machine learning approaches for predicting vessel integrity and reliability. For example, [7] developed a neural network-based model for predicting the remaining useful life of crude oil pressure vessels, while [8] used a decision tree-based approach to identify potential anomalies in vessel performance data. Despite being considered the safest means to transport oil and gas, pipelines are susceptible to degradation. Pipeline integrity management (PIM) is implemented to lower the risk of failure due to degradation and to maintain the functionality and safety of pipelines [13]. Pipelines are economical and efficient modes of transporting oil and gas. Pipelines will inevitably confront various risk factors throughout their lifespan, which could lead to defects. Defects in pipelines can compromise the integrity of the pipeline systems and may result in catastrophic accidents [14]. Corrosion is one of the many pipeline defects that mostly appear in a colony such that they interact to reduce the failure pressure, which is not defined by features of a single corrosion defect. The huge amount of corrosion defects captured by in-line inspection tools including the variability of defect profile in pipelines and the dependence of the reliability assessment on such data pose significant research challenges in performance assurance. In the research by [15], a novel approach is proposed for that involves computationally efficient modelling schemes to estimate the burst pressure of pipelines affected by both longitudinal and circumferential interacting corrosion defects by combining supervised machine learning methods with 25 numerical models of corroded pipelines, validated with experimental results available from literature. Catalytic cracking is a crucial process in petroleum refineries, enabling the conversion of heavy hydrocarbons into lighter, more valuable products [16]. This complex process is influenced by a multitude of variables, including temperature, pressure, catalyst properties, and feedstock characteristics [15]. Understanding the inter relationships between these variables is crucial for optimizing catalytic cracking performance, improving product yields, and reducing operating costs [17][18].

Numerous studies have investigated the effects of various variables on catalytic cracking performance. Temperature, for instance, has been identified as a critical factor, with optimal temperatures ranging from 480°C to 550°C [16]. Similarly, catalyst properties, such as surface area and pore size, have been found to significantly impact catalytic cracking performance [17] [18]. The impact of feedstock characteristics on catalytic cracking performance has also been extensively studied. For example, research has shown that feedstock density and viscosity can significantly affect product yields and selectivity [19] [20]. Additionally, the effects of operating conditions, such as pressure and residence time, on catalytic cracking performance have been investigated [21][22]. Multivariate analysis techniques, such as principal component analysis (PCA) and cluster analysis, have been employed to identify patterns and relationships between variables in catalytic cracking [20][21]. These studies have demonstrated the effectiveness of these techniques in reducing data dimensionality and identifying key variables that influence catalytic cracking performance. Machine learning algorithms, such as artificial neural networks (ANNs) and support vector machines (SVMs), have also been applied to model and optimize catalytic cracking processes [23][24]. These studies have shown that machine learning algorithms can be effective in predicting product yields and optimizing operating conditions. [25] combined expert knowledge and data analytics (Artificial Intelligence, Machine Learning, and Keyword Analysis) to create a reaction network for Asset Integrity Management (AIM) and provide a theoretical and practical basis for handling uncertainty in large data sets such as company incident databases. The purpose of their study was to control and minimize the total number of incidents that occur within an oil and gas operation by applying a multidisciplinary approach to explore and develop AIM. [26] evaluated the current state of the Bayesian network approach, which included methodology, influential parameters, and datasets for risk analysis, and to provide industry experts and academics with suggestions for future enhancements using content analysis. In the work by [27], modified IFM was utilized and by converting the pressure vessel failure design data to that of a wide tensile plate by changing its equivalent crack length obtained from an axial crack present in the cylinder. Results from the analysis carried out by [28] revealed that the incorporation of degradation and condition-based maintenance (CBM) can indeed be done and significantly influence the reliability analysis and planning of offshore energy assets.

Recent advances in sensing and computing technology have given rise to predictive maintenance techniques which, unlike traditional maintenance management techniques, attempt to predict failures and avoid system shut down proactively. The present study implements asset integrity management as pivotal for safe and efficient operation by employing Data-Driven Non-parametric Test and Machine Learning Approach for Crude Oil Pressure Vessels.

2. METHODOLOGY

A commercial oil and gas company located in Kwale, Delta State, Nigeria was selected as the model used in the study. In the development of questionnaires, ranking the variables that addresses the specific context of the ‘AIR’. Baseline understanding of current compliance levels is ensured in tailoring the questions to actual conditions and practices within the organization or industry. These practices were obtained from Inspection reports and compliance documents related to AIR, and survey-interviews with maintenance and inspection personnel. The questionnaires comprised of thirty-two

(32) variables for AIR administered to thirteen (13) experts for 'W' ranking and thirty-two (32) variables for NDT techniques administered to thirteen (13) experts for the Random Forest Classifier. In developing the questionnaire survey, the questions were clear and unambiguous, containing the required information in a standardized format that did not lead the respondent to more than one interpretation. The scaled ranking varied from 1 to 32, in which 1 is regarded as the most important and 32, the least important. Fig. 1 describes the outline of the methodology.

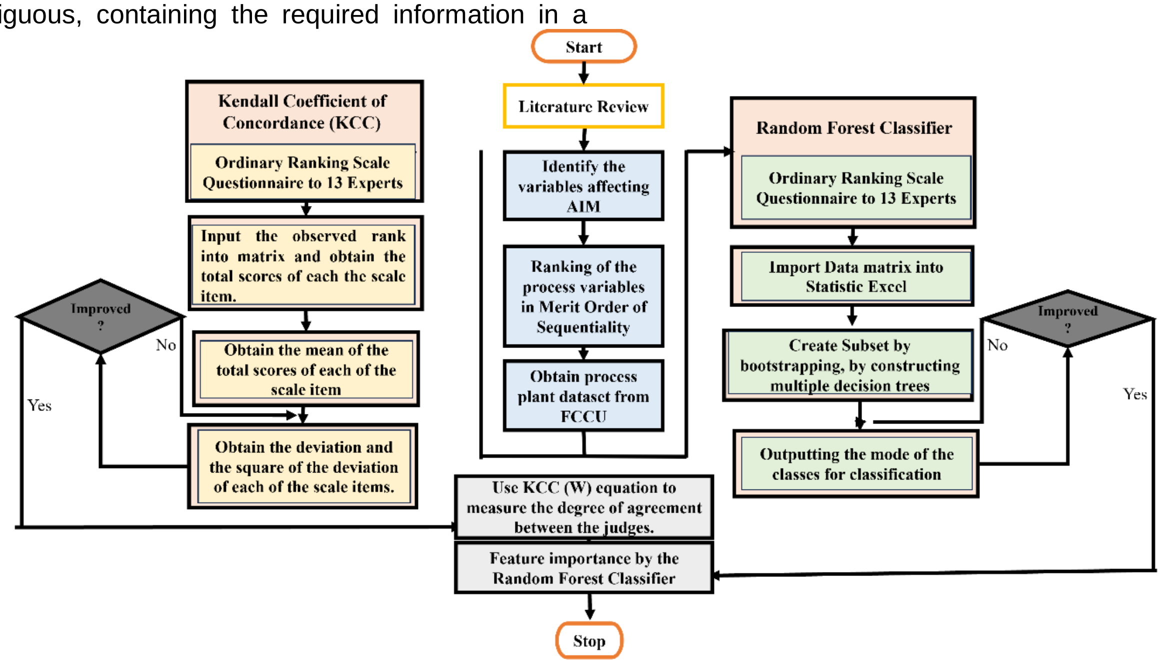

Fig. 1: Description of Methodology

The sequence of this analysis in Fig. 1, involves the data and mathematical analysis for the Kendall's Coefficient of Concordance, Merit order of variables' sequentiality and the Random Forest Classifier.

2.1 Kendall Coefficient of Concordance (W)

'W' is useful in establishing merit order sequence of the variables. The Kendall Coefficient of Concordance (W), which measures the degree of agreement between the judges is obtained from the Eqn. (1) and (2) [31].

$$W = \frac{S}{\frac{1}{12}K^2(N^3 - N)} \quad (1)$$

where $s = \sum\left(R_j - \frac{\sum R_j}{N}\right)^2$ = Rank variance (2)

Where R_j is the Column sum of ranks, N, the total number of Variables being ranked, S, the sum of Variance and K, the number of experts.

2.2 Random Forest Classifier

The Random Forest classifier, developed by [29], is an ensemble learning method that combines multiple decision trees to enhance prediction accuracy and robustness, using the formula for aggregating predictions through majority voting for classification. Random Forests have been proven to provide good results in cases where there are nonlinear relations between the variables in numerous industries [30]. The algorithm was chosen because of success in similar research and the properties of the technique with regards to overfitting, nonlinearity, and overall performance. Each subset is created by bootstrapping, that operates by constructing multiple decision trees during training and outputting the mode of the classes for classification or the mean prediction for regression Mathematically, the final prediction of the Random Forest *D(x)* is obtained by majority voting (classification) is shown in Eqn (3).

$$D(y) = \text{mode}\ \{h_i(y) \mid i = 1, 2, \ldots, k\} \quad (3)$$

Each tree P_i makes a prediction $h_i(y)$ for a given input *y.* The final prediction n of the Random Forest *D(y)* is obtained by majority voting (classification).

3.0 RESULT AND DISCUSSION

The Kendall Coefficient of Concordance, W, computation in Eqn (1) is 0.78. This value indicates that it is 'meritorious'. The significant values were obtained at 99.6 according to the test using the chi square table at

0.01 level of significance. This prompted rejecting the null hypothesis that disconcordance exists among the experts' rankings. Therefore, it is concluded that the experts ranked all scaled variables using similar standards. As a result, the analytical tool was used to organize the scaled variables according to their order of sequentiality, as shown in Table 1.

Table 1: 'W' ranked scaled variables regarding 'AIR'

S/N		Factors	S/N		Factors
1.	79	Corrosion	17.	211	Design Standards
2.	84	Pressure	18.	218	Life Cycle Management
3.	87	Temperature limits	19.	228	Quality Assurance
4.	93	Vibration	20.	237	Chemical Compatibility
5.	96	Maintenance Strategies	21.	254	Fluid Properties
6.	106	Inspection Techniques	22.	268	Data Management
7	139	Operational Procedures	23.	273	Training and Competence of operators
8	147	Regulatory Compliance	24.	279	Asset Documentation
9	151	Risk Assessment	25.	283	Root Cause Analysis
10	166	Failure Analysis	26.	288	Asset Optimization
11	169	Safety Instrumented Systems	27.	296	Equipment Monitoring
12	173	Emergency Response	28.	304	Safety Culture
13	179	Environmental Conditions	29.	318	Supply Chain Management
14	196	Human Factors	30.	332	Environmental Impact
15	198	Material Properties	31.	342	Asset Tracking
16.	208	Asset Monitoring	32	344	Regulatory Changes

3.2 Random Forest Classifier

The other set of questionnaires prioritized NDT techniques based on their effectiveness, precision, and broad applicability in the oil and gas industry's pressure vessel inspections was developed and administered to experts in the oil and gas industry. The experts ranked and prioritized 32 different advanced NDT techniques based on their effectiveness, precision, and broad applicability in the oil and gas industry's pressure vessel inspections. The selected features are temp. ctrl limits, environmental impact, asset document, failure analysis, equipment monitoring, vibration, pres. ctrl limits, regulatory compliance, human factors and safety culture. Hyperparameter tuning significantly enhances model performance and reliability. By carefully selecting and optimizing hyperparameters [37], the models are ensured to be both effective and efficient. Scikit-learn library [38] was applied for hyper-parameter optimization. The library for random search cross-validation was utilized to find the best hyperparameters for the Random Forest Classifier Model. Table 2 shows the RF model hyperparameter limits and optimized values, while Fig. 2 illustrates the ranked NDT techniques by the Random Forest Classifier.

Table 2: Models Hyperparameter limits and optimized values

Model	Hyperparameters	Range	Optimized Value
Random Forest	n_estimators	10 - 200	100
	max_depth	10 - 30	20
	min samples split	2 - 10	5
	min samples leaf	1 - 4	2

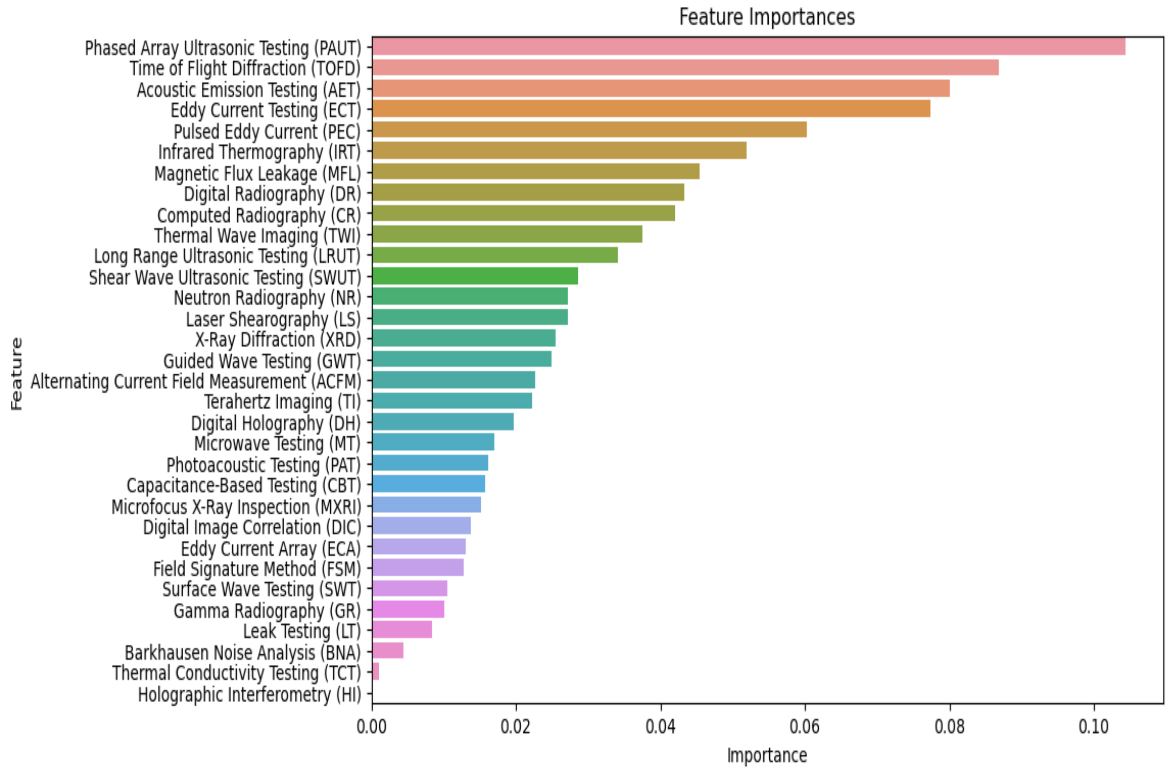

Fig. 2: Ranked NDT techniques by the Random Forest Classifier.

The horizontal bar chart in Fig. 2 displays the ranking of feature importance scores assigned by the Random Forest model. The features are listed on the y-axis, while their corresponding importance scores are shown on the x-axis. Phased Array Ultrasonic Testing (PAUT) is ranked as the most important feature, followed closely by Time-of-Flight Diffraction (TOFD), Acoustic Emission Testing (AET), and Eddy Current Testing (ECT). These top features have the highest contribution to the model's decision-making. Other significant features include Pulsed Eddy Current (PEC), Infrared Thermography (IRT), Magnetic Flux Leakage (MFL), and Digital Radiography (DR), which also demonstrate relatively high importance scores. Computed Radiography (CR), Thermal Wave Imaging (TWI), and various ultrasonic and radiographic testing methods contribute moderately to the model's predictions. As the ranking progresses downward, features such as Digital Holography (DH), Microwave Testing (MT), Terahertz Imaging (TI), and Capacitance-Based Testing (CBT) show lower importance. The least significant features include Barkhausen Noise Analysis (BNA), Thermal Conductivity Testing (TCT), and Holographic Interferometry (HI), which contribute minimally to the model. The ranking suggests that ultrasonic and eddy current-based methods play a crucial role in the model's classification, while techniques like holography and thermal conductivity have much lower influence.

5. CONCLUSION

In the advancement of the industry 4.0 vision over the years, significant changes and innovations brought about by analytic disruptions using advanced analytics and data-driven technologies have sharpened the oil and gas industry. These disruptions are transforming traditional business models, operations, and decision-making processes by leveraging big data, machine learning, artificial intelligence (AI), and other advanced analytics tools. The Oil and gas industry as a decisive element of the world's economy and energy reservoir consists of such different global processes as exploration, extraction, transporting, refining and marketing. Incidents in the oil and gas industry have caused financial loss, environmental damage, and broader societal concerns due to ineffective Process Safety Management (PSM). These techniques are prioritized based on their effectiveness, precision, and broad applicability in the oil and gas industry's pressure vessel inspections. The Kendall Coefficient of Concordance showed that Corrosion (79), Pressure (84), Temperature limits (87), Vibration (93), Maintenance Strategies (96), Inspection Techniques (106) are the most important 'AIR' parameters while the Phased Array ultrasonic testing, Time of flight Diffraction, Acoustic Emission Testing, Eddy current testing, Pulsed Eddy current are the most important NDT techniques. The study provides insights into which features are most influential as regards AIR and NDT techniques, highlighting the strong impact and weaker contributions from Expert opinion.

5. 0 Declaration of Competing Interest

The authors declare no conflict of interest.

REFERENCES

[1] Tang, K. H. D. (2021). A case study of asset integrity and process safety management of major oil and gas Companies in Malaysia. Journal of Engineering Research and Reports, 20(2), 6-19.

[2] Bouhala, L., Karatrantos, A., Reinhardt, H., Schramm, N., Akin, B., Rauscher, A., Mauersberger, A., Taşkıran, S.T., Ulaşlı, M.E., Aktaş, E. and Tanoglu, M., 2024. Advancement in the modeling and design of composite pressure vessels for hydrogen storage: A comprehensive review. Journal of Composites Science, 8(9), p.339.

[3] Ibrion, M., Paltrinieri, N., & Nejad, A. R. (2020). Learning from failures: Accidents of marine structures on Norwegian continental shelf over 40 years time period. Engineering Failure Analysis, 111, 104487.

[4] Xu, Z., & Saleh, J. H. (2021). Machine learning for reliability engineering and safety applications: Review of current status and future opportunities. Reliability Engineering & System Safety, 211, 107530.

[5] Obioha Val, O., Olaniyi, O. O., Selesi-Aina, O., Gbadebo, M. O., & Kolade, T. M. (2024). Machine Learning-enabled Smart Sensors for Real-time Industrial Monitoring: Revolutionizing Predictive Analytics and Decision-making in Diverse Sector. Machine Learning-enabled Smart Sensors for Real-time Industrial Monitoring: Revolutionizing Predictive Analytics and Decision-making in Diverse Sector (November 22, 2024). Asian Journal of Research in Computer Science, 17(11), 10-9734.

[6] Wang, H., Barone, G., & Smith, A. (2024). Current and future role of data fusion and machine learning in infrastructure health monitoring. Structure and Infrastructure Engineering, 20(12), 1853-1882.

[7] Kumar, A., et al. (2020). Predicting remaining useful life of crude oil pressure vessels using neural networks. Journal of Petroleum Science and Engineering, 195, 107324.

[8] Singh, H., et al. (2020). Anomaly detection in crude oil pressure vessel performance data using decision trees. Journal of Loss Prevention in the Process Industries, 63, 104021.

[9] Li, X., et al. (2019). A review of data-driven approaches for predictive maintenance in the oil and gas industry. Journal of Intelligent Manufacturing, 30(4), 1511-1525.

[10] Wang, J., et al. (2020). A review of machine learning applications in the oil and gas industry. Journal of Petroleum Science and Engineering, 186, 107522.

[11] Hollander, M., et al. (2018). Nonparametric statistical methods. John Wiley & Sons.

[12] Bishop, C. M., et al. (2016). Pattern recognition and machine learning. Springer.

[13] Rachman A, Zhang T, Ratnayake RC. Applications of machine learning in pipeline integrity management: A state-of-the-art review. International journal of pressure vessels and piping. 2021;193:104471.

[14] Ling, J., Feng, K., Wang, T., Liao, M., Yang, C. and Liu, Z., 2023. Data modeling techniques for pipeline integrity assessment: A state-of-the-art survey. IEEE Transactions on Instrumentation and Measurement, 72, pp.1-17.

[15] Mensah, A. and Sriramula, S., 2023. Estimation of burst pressure of pipelines with interacting corrosion clusters based on machine learning models. Journal of Loss Prevention in the Process Industries, 85, p.105176.

[16] Speight, J. G. (2017). Handbook of petroleum product analysis. John Wiley & Sons.

[17] Meyers, R. A. (2016). Handbook of petroleum refining processes. McGraw-Hill Education.

[18] Al-Mutairi, S. A., Al-Humaidan, F. A., & Al-Saleh, M. A. (2016). Optimization of catalytic cracking process using response surface methodology. Journal of Petroleum Science and Engineering, 147, 351-362.

[19] Zhang, Y., Li, X., & Liu, Y. (2020). Optimization of catalytic cracking process using machine learning algorithms. Fuel, 277, 118144.

[20] Kumar, P., Kumar, A., & Sharma, S. (2017). Application of principal component analysis in catalytic cracking unit. Journal of Process Control, 55, 125-134.

[21] Singh, J., Kumar, P., & Sharma, S. (2020). Cluster analysis and principal component analysis for identifying key variables in catalytic cracking unit. Journal of Process Control, 96, 103926.

[22] Liu, X., Liu, Y., & Li, X. (2019). Optimization of catalytic cracking process using artificial neural networks. Fuel, 235, 145-153.

[23] Wang, Y., Li, X., & Liu, Y. (2020). Optimization of catalytic cracking process using support vector machines. Journal of Petroleum Science and Engineering, 196, 107612.

[24] Liu, Y., Zhang, Y., & Li, X. (2020). Machine learning-based modeling and optimization of catalytic cracking process. Energy & Fuels, 34(5), 6513-6523.

[25] Sattari F, Lefsrud L, Kurian D, Macciotta R. A theoretical framework for data-driven artificial intelligence decision making for enhancing the asset integrity management system in the oil & gas sector. Journal of Loss Prevention in the Process Industries. 2022; 74:104648.

[26] Soomro, A.A., Mokhtar, A.A., Kurnia, J.C., Lashari, N., Sarwar, U., Jameel, S.M., Inayat, M. and Oladosu, T.L., 2022. A review on Bayesian modeling approach to quantify failure risk assessment of oil and gas pipelines due to corrosion. International Journal of Pressure Vessels and Piping, 200, p.104841.

[27] Agu, M.J., Gopikumar, S., Vimal, S. and Robinson, Y.H., 2020. Failure assessment of pressure vessels made of plain carbon steel by using modified inherent flaw model in DL based industry optimization intelligent processing. Measurement, 165, p.108112.

[28] Elusakin, T., 2021. Advanced reliability analysis of complex offshore Energy systems subject to condition based maintenance.

[29] M. Payette and G. Abdul-Nour, ‘Machine Learning Applications for Reliability Engineering: A Review’, Sustainability, vol. 15, no. 7, p. 6270, Apr. 2023, doi: 10.3390/su15076270.

[30] M. Abyani, M. R. Bahaari, M. Zarrin, and M. Nasseri, ‘Predicting failure pressure of the corroded offshore pipelines using an efficient finite element based algorithm and machine learning techniques’, Ocean Engineering, vol. 254, p. 111382, Jun. 2022, doi: 10.1016/j.oceaneng.2022.111382.

[31] O.J. Oyejide, A. Faiz, A. Muhammad, M.O. Okwu, Application of Decision Support Expert Systems for Improved gasoline yield in Refinery Catalytic Cracking, Procedia Comput Sci 232 (2024) 3044–3053. https://doi.org/10.1016/j.procs.2024.02.120.