# ENHANCEMENT OF MAINTENANCE STRATEGIES IN CRUDE OIL REFINERY PLANT EQUIPMENT USING K-NEAREST NEIGHBORS (KNN) MODEL

**Johnson Mbonu**[1]
Department of Electrical/Electronics Computer Engineering,
University of Uyo, Akwa Ibom, Nigeria

**Umana I. Itaketo**[2]
Department of Electrical/Electronics Computer Engineering,
University of Uyo, Akwa Ibom, Nigeria
uwanaitaketo@uniuyo.edu.ng

**Kufre M. Udofia**[3]
Department of Electrical/Electronics Computer Engineering,
University of Uyo, Akwa Ibom, Nigeria
kufremudofia@uniuyo.edu.ng

***Abstract*— In this work, enhancement of maintenance strategies in crude oil refinery plant equipment using k-Nearest Neighbors (KNN) model is presented. The study focused on predicting equipment failure and thereby optimizing maintenance strategies in the crude oil refinery plant. The KNN model is used for the prediction of the likelihood of equipment failure thereby triggering warning for preventive maintenance scheduling. The analysis is based exclusively on the available time-series data for discharge pressure, temperature, and vibration, spanning from early January 2025 to early April 2025. The original data contains only three parameters, pressure status, temperature status, and vibration status. The equipment (target) status is derived from the three available parameters. The dataset had 500 data records and the KNN model was trained with 75% of the data while validation evaluation was done with the remaining 25 % of the data. The results show that the KNN model has mean prediction accuracy of 87.75 %, precision of 91%, recall of 86 % and F1-score of 88% with support score of 253. Also, the equipment had 100 critical status before the application of the KNN model however after the application of the KNN model the critical status reduced to 31 instances which is about 69 % improvement in the maintenance scheduling mechanism.**

***Keywords*— *Preventive maintenance, Crude Oil Refinery Plant, Reactive Maintenance, k-Nearest Neighbors (KNN), Sensor Data***

## 1. INTRODUCTION

The oil and gas industry in Nigeria faces major hurdles in managing and maintaining its equipment, which is crucial for ensuring operational efficiency, safety, and reliability [1,2,3] The sophistication and scale of equipment in the oil and gas industry, coupled with rough operating environments, make traditional maintenance practices often ineffective and costly [4,5,6]. Conventionally, maintenance in refinery plants has relied heavily on reactive strategy in response to equipment failure and preventive strategy which seeks to avoid equipment failure [7,8]. Preventive maintenance, while more proactive, is scheduled at regular intervals regardless of the actual condition of the equipment, which can lead to unnecessary maintenance actions or failure to detect impending issues [9,10,11]. Both approaches have limitations in terms of cost efficiency, resource allocation and effectiveness in preventing critical failures.

Notably, in recent years, advancements in machine learning (ML) and data analytics have opened new opportunities for predictive maintenance, which aims to forecast equipment failures before they occur based on historical and real time operational data [14,15]. Predictive strategies basically utilize data from the equipment components status collected over time which are then fed to powerful algorithms to predict likelihood of impending fault [16,17]. When integrated into maintenance strategies, these techniques not only enhance equipment reliability and availability but also optimize maintenance schedules, reduce operational costs and improve safety performance [16]. Accordingly, the purpose of this work is to explore the role of AI-driven predictive maintenance in transforming equipment management in the oil and gas industry. The study employed the K-Nearest Neighbor (KNN) model on a sensor dataset obtained from a case study refinery. In all, this work aims to provide insights into how AI can revolutionize equipment management in the oil and gas industry, thereby improving efficiency, safety, and reliability across the value chain.

## 2. METHODOLOGY

The study is focused on predicting equipment failure and thereby optimizing maintenance strategies in the crude oil refinery plant. The K-Nearest Neighbor (KNN) model is used for the prediction of likelihood of equipment failure thereby triggering warning for preventive maintenance. The analysis is based exclusively on the provided time-series data for discharge pressure, temperature, and vibration, spanning from early January 2025 to early April 2025. The summary of the statistical description of the study dataset is shown in Table 1.

The original data contains only three parameters, pressure status, temperature status, and vibration status. The equipment (target) status is obtained from the three available parameters. The criteria used to determine the equipment (target) status from the three available parameters are presented in Table 2.

The scatter plot of the data items are shown in Figure 1 to Figure 4, Pressure Status (Figure 1), Temperature Status (Figure 2), Vibration Status (Figure 3) and Target Status (Figure 4). The pie chart for the plant (target) status showing the distribution of the three status in the dataset is presented in Figure 5. The summary of the number of incidence of the three status categories for each of the four parameters is presented in Figure 6. The data in Figure 5 and Figure 6 shows that there 100 instances of critical state status for the equipment which will result reactive maintenance for about 100 times with the period listed in the data collection.

Based on the dataset consisting of 500 data records, the KNN model was trained with 75% of the data and then validation evaluation was done with the remaining 25 % of the data. The procedure used to train and deploy the KNN model for the preventive maintenance in is presented as Algorithm 1.

**Table 1** The summary of the statistical description of the study dataset

| Groups | Pressure_Status | Temperature_Status | Vibration_Status | Target |
|---|---|---|---|---|
| Num of observations | 500 | 500 | 500 | 500 |
| Num of missing values | 0 | 0 | 0 | 0 |
| Minimum | 1 | 1 | 1 | 1 |
| Maximum | 3 | 3 | 3 | 3 |
| Range | 2 | 2 | 2 | 2 |
| Mean ($\bar{x}$) | 1.47 | 1.476 | 1.434 | 1.8 |
| Sum | 735 | 738 | 717 | 900 |
| Mean Confidence Interval , 95% CI | [1.4116, 1.5284] | [1.4165, 1.5355] | [1.3774, 1.4906] | [1.7342, 1.8658] |
| Standard Deviation (S) | 0.6648 | 0.677 | 0.6438 | 0.7491 |
| Q1 | 1 | 1 | 1 | 1 |
| Median | 1 | 1 | 1 | 2 |
| Q3 | 2 | 2 | 2 | 2 |
| Interquartile range | 1 | 1 | 1 | 1 |
| Skewness | 1.0964 | 1.0989 | 1.1978 | 0.3447 |
| Outliers | | | | |

**Table 2 The criteria used to determine the equipment (target) status from the three available parameters**

| Threshold Level | Condition | Example | Recommended Action |
|---|---|---|---|
| **Normal** | All sensor readings within normal range | Pressure = 45-65 psi, Vibration = 0.2g | Continue regular operation, log data for ML training |
| **Normal** | No threshold flags triggered | - | No action needed |
| **Warning** | One or more warning thresholds exceeded | Vibration = 0.5g (ISO limit > 0.45g) | Increase monitoring frequency |
| **Warning** | Gradual increase toward threshold limits | Temperature trending toward 260°C, Pressure = 68psi | Schedule inspection during next planned maintenance |
| **Warning** | ML model risk score in medium range (0.3–0.6) | - | Alert maintenance planner |
| **Critical** | Two critical threshold breached | Pressure > 68 psi, Temp > 280°C | Trigger emergency maintenance alert |
| **Critical** | Rapid abnormal trend | Vibration spike from 0.5g-0.7g | Trigger emergency shutdown and blow down |
| **Critical** | Multiple threshold breaches or ML risk score ≥ 0.6 | Pressure + Temp + Vibration out of range | Notify reliability team, perform forced inspection |

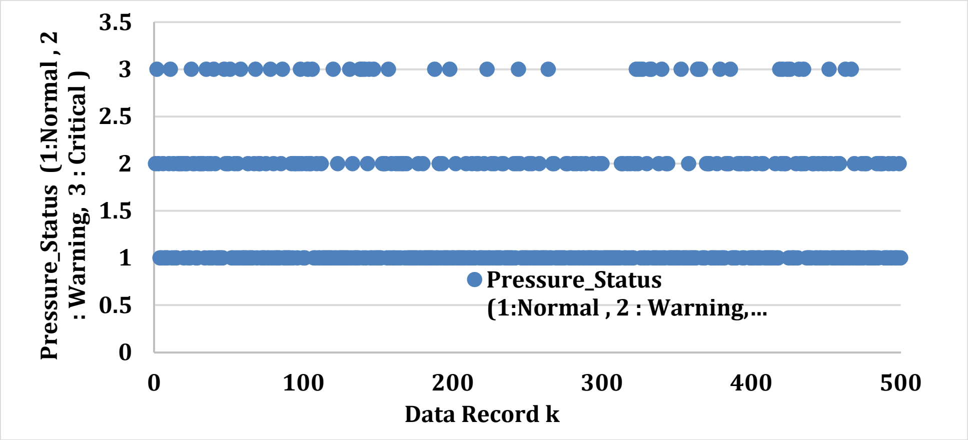



Figure 1 The scatter plot of the Pressure Status data

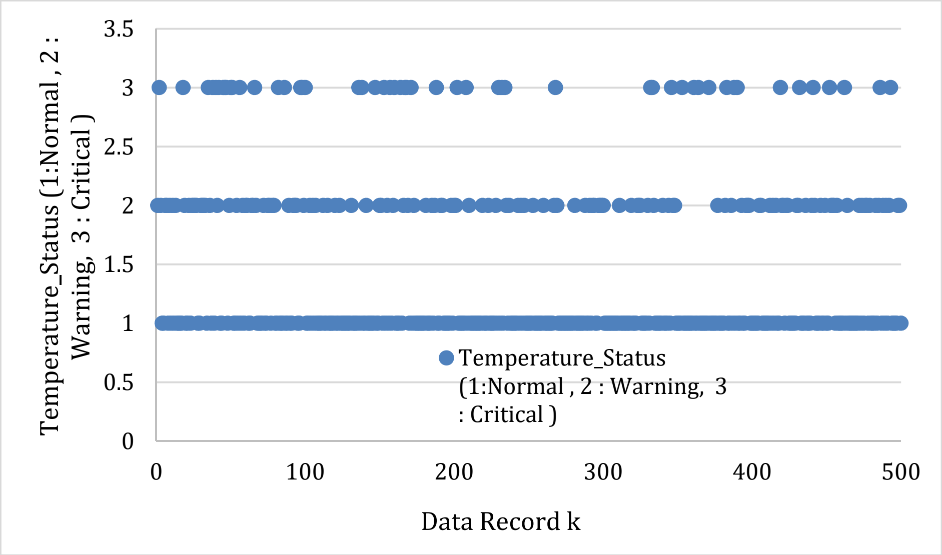



Figure 2 The scatter plot of the Temperature Status data

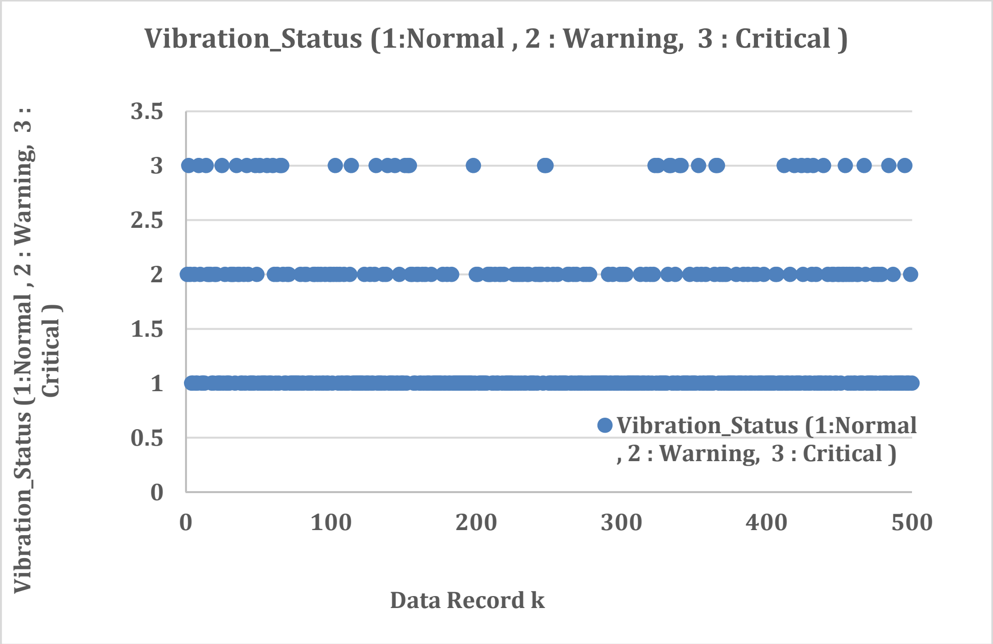



Figure 3 The scatter plot of the Vibration Status data

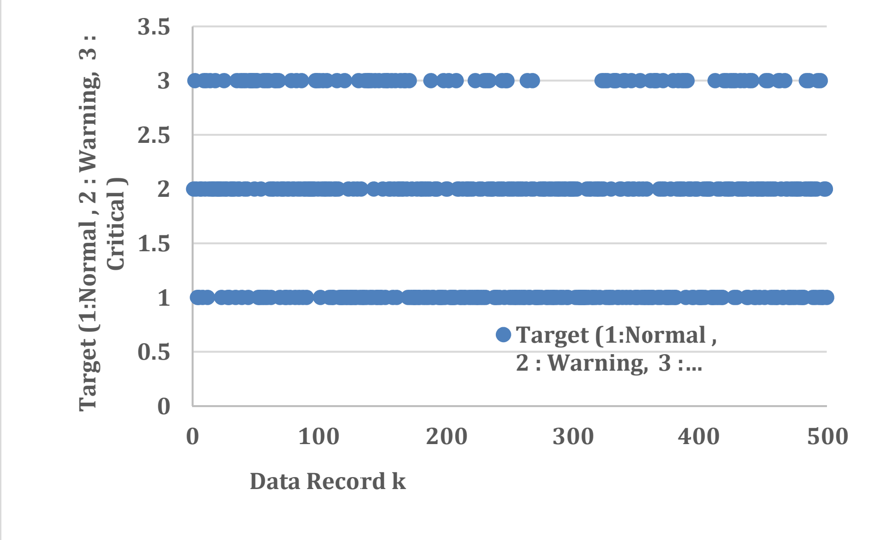



Figure 4 The scatter plot of the Vibration Status data

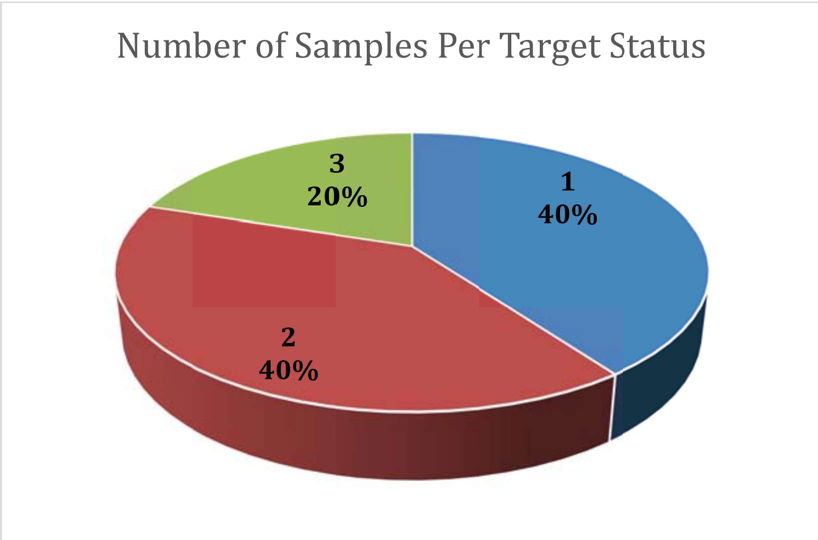
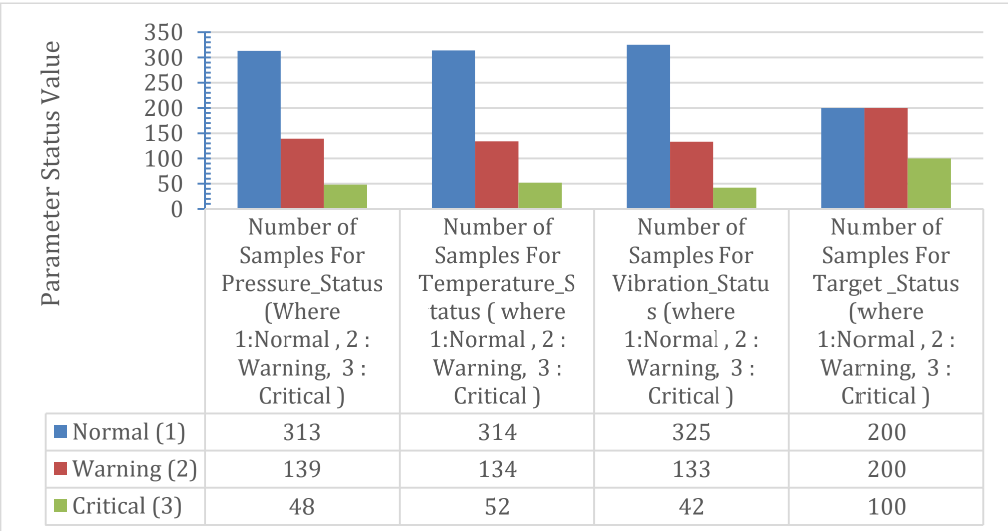



Figure 5 The pie chart for the plant (target) status showing the distribution of the three status in the dataset




| | Number of Samples For Pressure_Status (Where 1:Normal , 2 : Warning, 3 : Critical ) | Number of Samples For Temperature_Status ( where 1:Normal , 2 : Warning, 3 : Critical ) | Number of Samples For Vibration_Status (where 1:Normal , 2 : Warning, 3 : Critical ) | Number of Samples For Target _Status (where 1:Normal , 2 : Warning, 3 : Critical ) |
|---|---|---|---|---|
| Normal (1) | 313 | 314 | 325 | 200 |
| Warning (2) | 139 | 134 | 133 | 200 |
| Critical (3) | 48 | 52 | 42 | 100 |

**Figure 6 The summary of the number of incidence of the three status categories for each of the four parameters**

**Algorithm 1**

Step 1: Data Collection
    Step 1.1 Gather historical maintenance and operational data.
    Step 1.2 Collect sensor data from equipment.
Step 2 : Data Preprocessing:
    Step 2.1 Clean and preprocess the data (handling missing values, normalization).
    Step 2.1 Feature selection and engineering.
Step 3 Model Development:
    Step 3.1 Split the data into training and testing sets.
    Step 3.2 Train machine learning models (KNN).
    Step 3.3 Model Evaluation:
    Step 3.4 Validate model performance using metrics such as accuracy, precision, recall, and F1 score.
    Step 3.5 Adjust hyper-parameters for optimization.
Step 4 Prediction and Analysis:
    Step 4.1 Use the trained models to predict equipment failures.
    Step 4.2 Analyze the predicted outcomes to identify patterns and trends.
Step 5 Optimization of Maintenance Strategies:
    Step 5.1 Develop maintenance schedules based on predictive results.
    Step 5.2 Create a decision-making framework for proactive maintenance.
Step 6 Implementation and Monitoring:
    Step 6.1 Implement optimized maintenance strategies in the refinery.
    Step 6.16 Continuously monitor model performance and refine as necessary.

## 3. RESULTS AND DISCUSSION

### 3.1 Results for the KNN Model Training and Evaluation

The result show that the KNN model has prediction accuracy of mean 87.75 %, precision of 91%, recall of 86 % and F1-score of 88% with support score of 253, as shown in Table 3.

**Table 3: The KNN prediction performance**

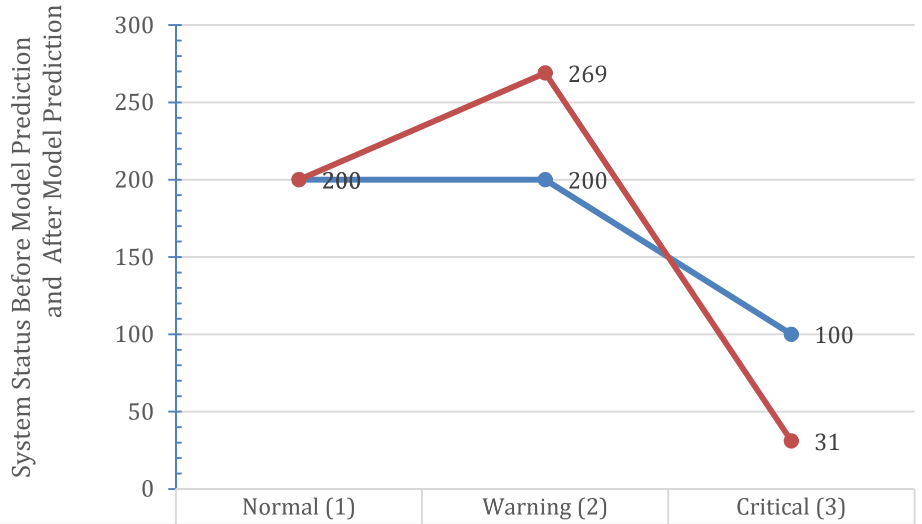
| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 1 (Normal) | 0.83 | 0.94 | 0.88 | 108 |
| 2 (Warning) | 0.91 | 0.83 | 0.87 | 123 |
| 3 (Critical) | 1.00 | 0.82 | 0.90 | 22 |
| Overall Accuracy | — | — | 0.88 | 253 |
| Macro Average | 0.91 | 0.86 | 0.88 | 253 |
| Weighted Average | 0.88 | 0.88 | 0.88 | 253 |

For the Class 1 (Normal) the KNN had high recall (0.94) which shows that the model successfully identifies most of the normal cases. Also, slightly lower precision (0.83) means some warnings/critical were predicted as normal. For the Class 2 (Warning) the KNN had very good balance between precision (0.91) and recall (0.83). Suggests the model understands this intermediate class quite well. For the Class 3 (Critical) the KNN had perfect precision (1.00). The recall (0.82)) means the KNN model misses a few critical cases, but still captures most. F1-score (0.90) reflects strong but improvable sensitivity to critical events.

In general, the KNN had very good prediction performance. The K-Nearest Neighbors (KNN) model also shows strong overall performance with accuracy at 87.75%. Moreover, the KNN demonstrated good precision, recall, and F1-scores across all classes, with macro and weighted averages generally in the range of 0.86 to 0.91.

After the prediction of equipment failure using KNN Model, the model was trained to flag warning signal status when two of the parameters records warning (2) and critical (3). With this, the reactive maintenance of the equipment and corrective maintenance is reduced by a preventive maintenance whereby the equipment will be booked for maintenance. With this, there will be reduction in downtime in the refinery plant.

It is clearly seen that before the predictive maintenance strategy, that the number of critical entries( breakdown) were 100 and after the model has been trained for the prediction of equipment failure, and was used to optimize the maintenance strategies, the critical entries (breakdown) reduced to 31 entries which has a significant impact on the equipment health. The summary of the number of incidence of the three status categories for equipment (target) status before the application of the KNN model and after the KNN model application is presented in Figure 7. It showed that the equipment had 100 critical status before the application of the KNN model however after the application of the KNN model the critical status reduced to 31 instances which is about 69 % improvement in the maintenance scheduling mechanism.




| | Normal (1) | Warning (2) | Critical (3) |
|---|---|---|---|
| Number of Samples Before model Prediction | 200 | 200 | 100 |
| Number of Samples After Model Prediction | 200 | 269 | 31 |

Figure 7 The summary of the number of incidence of the three status categories for equipment (target) status before the application of the KNN model and after the KNN model application

## 4. CONCLUSION

This study explored the application of machine learning (ML) for predictive maintenance in crude oil refinery plants, focusing on critical equipment such as compressors. The work aimed to predict equipment failures, optimize maintenance strategies, using sensor data (temperature, pressure, vibration). The k-Nearest Neighbors (KNN) was used and the results showed that the equipment failure was reduced by 69 % through the prediction of likelihood of equipment failure based on the sensor parameters. The study successfully demonstrated that ML-driven predictive maintenance can transform refinery operations by reducing unplanned downtime through early failure detection.

## REFERENCES

1. Nwulu, E. O., Elete, T. Y., Erhueh, O. V., Akano, O. A., & Omomo, K. O. (2024). Leveraging predictive modelling to enhance equipment reliability: A generic approach for the oil and gas industry. *International Journal of Engineering Research and Development*, *20*(11), 951-969.
2. Adebisi, B., Aigbedion, E., Ayorinde, O. B., & Onukwulu, E. C. (2021). A conceptual model for predictive asset integrity management using data analytics to enhance maintenance and reliability in oil & gas operations. *International Journal of Multidisciplinary Research and Growth Evaluation*, *2*(1), 534-54.
3. Arinze, C. A., Izionworu, V. O., Isong, D., Daudu, C. D., & Adefemi, A. (2024). Integrating artificial intelligence into engineering processes for improved efficiency and safety in oil and gas operations. *Open Access Research Journal of Engineering and Technology*, *6*(1), 39-51.
4. Adebisi, B., Aigbedion, E., Ayorinde, O. B., & Onukwulu, E. C. (2021). A conceptual model for predictive asset integrity management using data analytics to enhance maintenance and reliability in oil & gas operations. *International Journal of Multidisciplinary Research and Growth Evaluation*, *2*(1), 534-54.
5. Mohamed Almazrouei, S., Dweiri, F., Aydin, R., & Alnaqbi, A. (2023). A review on the advancements and challenges of artificial intelligence based models for predictive maintenance of water injection pumps in the oil and gas industry. *SN Applied Sciences*, *5*(12), 391.
6. Wanasinghe, T. R., Trinh, T., Nguyen, T., Gosine, R. G., James, L. A., & Warrian, P. J. (2021). Human centric digital transformation and operator 4.0 for the oil and gas industry. *Ieee Access*, *9*, 113270-113291.
7. Afolalu, S. A., Monye, S. I., Ogundipe, A. T., Azeez, T. M., & Ikumapayi, O. M. (2024, April). Maintenance Strategies and Procedures: an Overview. In *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)* (pp. 1-8). IEEE.
8. Onyeke, F. O., Odujobi, O., Adikwu, F. E., & Elete, T. Y. (2023). Revolutionizing process alarm management in refinery operations: Strategies for reducing operational risks and improving system

reliability. *Magna Scientia Advanced Research and Reviews*, *9*(2), 187-194.

9. Yazdi, M. (2024). Maintenance strategies and optimization techniques. In *Advances in computational mathematics for industrial system reliability and maintainability* (pp. 43-58). Cham: Springer Nature Switzerland.
10. Kalaiselvi, K., Niranjana, K., Prithivirajan, V., Kumar, K. S., & Syambabu, V. (2024, August). Machine Learning for Predictive Maintenance in Industrial Equipment: Challenges and Application. In *2024 4th Asian Conference on Innovation in Technology (ASIANCON)* (pp. 1-6). IEEE.
11. Vajpayee, A., Mohan, R., Gangarapu, S., & Chilukoori, V. V. R. (2024). REALTIME DATA PROCESSING IN PREDICTIVE MAINTENANCE: ENHANCING INDUSTRIAL EFFICIENCY AND EQUIPMENT LONGEVITY. *INTERNATIONAL JOURNAL OF ENGINEERING AND TECHNOLOGY RESEARCH (IJETR)*, *9*(2), 29-42.
12. Nwulu, E. O., Elete, T. Y., Erhueh, O. V., Akano, O. A., & Omomo, K. O. (2023). Machine learning applications in predictive maintenance: Enhancing efficiency across the oil and gas industry. *International Journal of Engineering Research Updates*, *5*(1), 17-30.
13. Virginia, E. O., Obada, O. R., Oke, B. O., Jimso, I. O., Iwalokun, A. O., Akinbolusere, M. O., & Pius, A. B. (2025). Leveraging machine learning and data analytics for equipment reliability in oil and gas using predictive maintenance. *World Journal of Advanced Research and Reviews*, *25*(1), 2212-2218.
14. Atassi, R., & Alhosban, F. (2023). Predictive Maintenance in IoT: Early Fault Detection and Failure Prediction in Industrial Equipment. *Journal of Intelligent Systems & Internet of Things*, *9*(2).
15. Geng, S., & Wang, X. (2022). Predictive maintenance scheduling for multiple power equipment based on data-driven fault prediction. *Computers & Industrial Engineering*, *164*, 107898.
16. Ewin, G., & Oye, E. (2025). Transitioning from Corrective to Preventive Maintenance Strategies: Enhancing Equipment Reliability and Efficiency.